

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: January 22, 2002, 16:23:40 ; Search time 40.57 Seconds  
(without alignments)  
1137.482 Million cell updates/sec

Title: US-09-485-529-7  
Perfect score: 3194  
Sequence: 1 MKREYQDAGSGGGGGGMS.....TLGWHTPLTATSAWRLAGP 623

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_1101.\*  
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3194	100.0	623	20	AAV02540
2	2683	84.0	630	20	AAV02541
3	2189	68.5	425	20	AAV02539
4	1612	50.5	587	22	AAE02560
5	1612	50.5	587	22	AAE01892
6	1594	48.9	532	18	AAW30792
7	1594	48.9	533	21	AAW38575
8	1590	49.8	532	22	AAE02545
9	1590	49.8	532	22	AAE01907
10	1589	49.7	517	21	AAW38577
11	1589	49.7	518	21	AAW38576

12	1588.5	49.7	531	21	AAW28575	Arabidopsis SCla2.
13	1572	49.2	572	22	AAW31883	Amino acid sequenc
14	1568	49.1	572	22	AAW31884	Amino acid sequenc
15	1542	47.7	587	21	AAW28574	Arabidopsis SCla8.
16	883.5	27.7	256	20	AAV02538	protein encoded by
17	812.5	25.4	277	18	AAW38193	Arabidopsis SCARC
18	797.5	25.0	262	18	AAW38194	Arabidopsis SCARC
19	673.5	21.1	138	20	AAV02544	protein encoded by
20	670	21.0	668	21	AAW28595	Maize Scarecrow pr
21	600.5	18.8	653	18	AAW38178	Arabidopsis SCARC
22	600.5	18.8	653	21	AAW28569	Arabidopsis SCARC
23	599	18.5	428	19	AAW81753	Tomato Ls protein.
24	589.5	18.5	282	18	AAW30795	Arabidopsis thalia
25	569.5	17.8	259	18	AAW30794	Arabidopsis thalia
26	554	17.3	384	21	AAW29710	Arabidopsis thalia
27	554	17.3	405	21	AAW29709	Arabidopsis thalia
28	554	17.3	413	21	AAW29708	Arabidopsis thalia
29	523.5	16.4	405	21	AAW48516	Arabidopsis thalia
30	513	16.1	313	21	AAW28577	Maize SC1m1. Zea
31	512	16.0	313	18	AAW38209	Maize Zm-SC12 prot
32	512	16.0	313	21	AAW28573	Maize SC1m1. Zea
33	507	15.9	306	18	AAW38179	Arabidopsis SCARC
34	507	15.9	306	21	AAW28570	Arabidopsis SCla4.
35	506	15.8	306	21	AAW28598	protein encoded by
36	498	15.6	480	21	AAW21049	Arabidopsis thalia
37	498	15.6	578	21	AAW21048	Arabidopsis thalia
38	488	15.6	593	21	AAW21047	Arabidopsis thalia
39	464	14.5	352	21	AAW28596	Arabidopsis thalia
40	462.5	14.5	323	21	AAW28603	protein encoded by
41	458	14.3	352	18	AAW38181	Maize CBPPT4 and
42	448	14.0	113	21	AAW32809	Arabidopsis SCARC.
43	443	13.9	809	21	AAW28586	Eucalyptus grandis
44	441	13.8	808	21	AAW28602	Arabidopsis SCla2
45	439	13.7	541	21	AAW50885	protein encoded by

#### ALIGNMENTS

RESULT 1	AAV02540	standard; protein; 623 AA.
ID	AAV02540	
XX	AAV02540:	
AC	16-JUL-1999	(first entry)
XX		
DT		
XX		
DE		Protein encoded by wheat Rht clone 5a1 genomic sequence.
XX		
KW	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;	
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;	
KW	paclobutrazol.	
XX		
OS	Triticum aestivum.	
XX		
PN	WO9909174-A1.	
XX		
PD	25-FEB-1999.	
XX		
PF	07-AUG-1998:	98WO-GB02383.
XX		
PR	13-AUG-1997:	97GB-0017192.
XX		
PA	(PLAN-) PLANT BIOSCIENCE LTD.	
XX		
PI	Harberd NP, Peng J, Richards DE;	
XX		
DR	WPI, 1999-181040/15.	
XX	N-PSDB: AAX36279.	
XX		
PT	New Triticum aestivum polynucleotides - encode a polypeptide which	
PT	provides inhibition of the growth of plants, which inhibition is	
PT	antagonised by gibberellin, used to confer a dwarf phenotype	

XX Disclosure: Fig 8b; 88pp; English.

XX The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in *Triticum aestivum*, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence is encoded by  
CC the wheat Rht clone 5a1 genomic sequence.

XX  
XX  
XX Sequence 623 AA:

Query Match 100.0%; Score 3194; DB 20; Length 623;  
Best Local Similarity 100.0%; Pred. No. 4,4e-244;  
Matches 623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKREYODAGSGGGGGGSGGSSSEDKMVAASAAGEEVEDLALALGYKVRASDMADVAQKL 60  
Db 1 mkreydgagsggggggsggmsedkmmvsaagaageevdellaalgykvrasmadvagkl 60

QY 61 EQLEAMKMGVYGAGAAADDDSFATHLADTYHYNFTDLSWWSWMSLSELMAPPPLPRAP 120  
Db 61 eqleammgvgyagaapddsfatlactdyhyntfclsswswmslseinappplppap 120

QY 121 QUNASTSTVYSGSGYFPLPPEVDSSSTLYALRPPSPAGATAPADLSADSVDRDKRMRT 180  
Db 121 qlnaststvgsgyfdlppsavdssstlyalrppspagatapadlsadvdrdkrmt 180

QY 181 GGSSTSSSSSSSSSLGGGARSSVVEAAPVVAANAATPALPVVVVDTOEGATRLVHALLA 240  
Db 181 ggsstssssssslgggarssvveaapvvaanaatpalpvvvdtgeagtrlvhalla 240

QY 241 CAEAVQOENLSAAELVYQIPLLAASOGGAMKVAAYTGEALARRVFRPRPDBSSLDA 300  
Db 241 caeavqgenlsaaelvkqipllaasgggamtkaaytgealarrvfrfppdbssllda 300

QY 301 AFADLLAHFHYSCPYLKFAHFTANQATLEAFAGCRRYVVDGFKOGQMOPALLOALAL 360  
Db 301 afadllahfyescpylkfahftanqalleafagcrryhvvdfgkqgmopalqalal 360

QY 361 RRGPPSRLTGVGPPDPDETDAQQVGVKTLAQFAHTTRVDFQYRGVLVAATLADLEPML 420  
Db 361 rpgppsrfltgvgpppddetdaqqvgvkwklaqfahtrrvdfqyrgvlvaatladlepml 420

QY 421 QPEGEDDNEPEEYVAVSVFEMHLLAQPGALDEVLGTVRAVRRTYVYVQENHNHSG 480  
Db 421 qpegeddnepeevlavsvfemhllaqpgaldevlgvtravrrtylvveqeanhns 480.

QY 481 TFLUDFTESLHYSTMPDSELEGSSGGGSEVSSGAAAPAAAGTDQVWSEYLLROICN 540  
Db 481 tfludfteeslhystmppslegssgggsevs sgaaapaaagtdqvwseyllrtqcn 540

QY 541 VVACGAERTERHETLIGWRNRLGNAGETVHALGSNAYKQASTLLALFAGDGYVEEKE 600  
Db 541 vvacegaertethelgqwrnrlgnagetvhlgnaykqaastllalfagdgylveeke 600

QY 601 GCLTLGWHTRPLIANTSAMRLACP 623  
Db 601 gcltlgwhtrpliatlsawrlagp 623

RESULT 2  
AAV02541  
ID AAV02541 standard; Protein: 630 AA.

XX	AA02541;
AC	
XX	16-JUL-1999 (first entry)
DT	
XX	
DE	Protein encoded by maize 1a1 genomic clone sequence.
XX	
KW	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
KM	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW	paclobutrazol; maize.
XX	
OS	Zea mays.
XX	
PN	MO909174-A1.
PD	
PD	25-FEB-1999.
XX	
PF	07-AUG-1998; 98MO-G802383.
XX	
PR	13-AUG-1997; 97GB-0017192.
XX	
PA	(PLAN-) PLANT BIOSCIENCE LTD.
XX	
PI	Harberd NP, Peng J, Richards DE;
XX	
DR	WPI: 1999-181040/15.
N-PSDB:	MAX36280.
XX	
PT	New Triticum Aestivum polynucleotides - encode a polypeptide which
PT	provides inhibition of the growth of plants, which inhibition is
XX	antagonised by gibberellin, used to confer a dwarf phenotype
XX	
PS	Disclosure; Fig 9b; 88pp; English.
CC	
CC	The specification describes polypeptides encoded by the Rht gene (and
CC	its homologues) that, when expressed in Triticum Aestivum, inhibit
CC	growth of the plant. This growth inhibition is antagonised by
CC	gibberellin. The products can be used to provide Rht expression in
CC	plants, conferring a dwarf phenotype on a plant which is correctable
CC	by treatment with gibberellin. In addition, the products can be
CC	used to produce Rht mutant plants which are dwarfed compared with
CC	wild-type, the dwarfing being gibberellin-insensitive. Taller plants
CC	may be made by knocking out Rht or the relevant homologous gene in
CC	the plant of interest. Plants may be made which are resistant to
CC	compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
CC	e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
CC	dwarf but let crop plants grow tall. The present sequence is encoded by
CC	the maize 1a1 genomic clone sequence.
XX	
SO	Sequence 630 AA;
Query Match	84.0%; Score 2683; DB 20; Length 630;
Best Local Similarity	85.8%; Pred. No. 1.le-203;
Matches 546; Conservative	27; Mismatches 37; Indels 26; Gaps 13;
QY	1 MKREYDAGSGGGGGSSEDKKMWSA--AGEGE--VDELALGKYVRASDMADVA 57
DB	1 mkreydgags---ggdmgsakdkmmaaaagaageeedvdellaalykrssdmadva 57
QY	58 OKLEDLBNAMNGCV-GGAAPDDSFATNLTDTVHYHPTLDSNVEMLSFINARPPRL 116
DB	58 qleqldemamngvgvgaagatddfvshlactdthyprsdslsvwemselineapparl 117
QY	117 P---PAPOLNASTSTVY----GSGGYEDLPSPVSSSSYALRPFPACATAPADLSA 169
DB	118 ppatcpaprl-asstscvtsgaaaagayridlpavovssscyalpripvaaps-adpst 175
QY	170 DSVRPMKMRTPGSSSTSSSSSSSLGCG-ARRSVEAAPR--VAAAANPALPVVVV 225
DB	176 dsarcpkmtirgyggtsssssssmgdgrttrsvveeaaapatcgaasaang-pavrvvv 234
QY	226 DTGEGIRLVHALLACAAVDOENISAAELVKQIPPLLAASOGCAMRVVAAYFGEALLAR 285

```

|||||
Db 235 dtgeagirlvhalldacaeavqgenfsaaealvkqjpmlasssggmkrkvaayfgealarr 294
QY 266 VFRFRPODSSLLDAARFADLHAHFYESCPIYKFAHFTANQAILAEAFACCRVHVVDGCI 345
Db 295 YRFRPPDSSLLDAARFADLHAHFYESCPIYKFAHFTANQAILAEAFACCRVHVVDGCI 354
QY 346 KOGMOWPALLQALARPGGPSPFRITGVGPPDPDFTDALQOYGMWLQAPAHRTIRVDFQYR 405
Db 355 Kqgmwppallqalarrpgppspfriltvgpppdeudalqvgwklagfahltirvdfqyr 414
QY 406 GLVAAATLADLEPFMLQPGSEEDPNEPEVIAVNSVFEMHRLAQGALAEKVLGTVRAVAP 465
Db 415 glvaatladdlpfmqpgg-ddtddepevianvsfclhrllagpalekvlgtvrvavp 473
QY 466 RIVVVEODPANHNSGFTLDRFTESLHYSTWFDLSLEGSSGGSPSEVSSGAAPAACT 525
Db 474 rlvvvegeanhnsgftldrfeslhystmfdslagagagsgstdaspa----aagtc 529
QY 526 DOVMSEVYLGROICNVVACGEAERTERHETTGOMNRRLGNAGFEVYHGSNMYKQASTLL 585
Db 530 dqvmsevylygrqicnvacegaertelcgrwrsrlgsgfapvhlgsnaykqastll 589
QY 586 ALFAGDGYKVEKEGCTLTGWHTRPLIATSAMRLA 621
Db 590 alfagdgryveekdgcltlgwhtrpliatssarva 625

RESULT 3
AA02539
ID AAY02539 standard; Protein: 425 AA.
XX
AC AAY02539;
XX
DT 16-JUL-1999 (first entry)
XX
DE Protein encoded by the wheat Rht clone C15-1 cDNA sequence.
XX
KW Rht gene; homologue: Triticum aestivum; wheat; growth inhibition;
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW paclobutrazol.
XX
OS Triticum aestivum.
XX
PN W09909174-A1.
XX
PD 25-FEB-1999.
XX
PE 07-AUG-1998; 98WO-GB02383.
XX
PR 13-AUG-1997; 97GB-0017192.
XX
PA (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI Harberd NP, Peng J, Richards DE;
XX
DR WPI: 1999-181040/15.
DR N-PSDB: MAX36278.
XX
PT New Triticum Aestivum polynucleotides - encode a polypeptide which
PT provides inhibition of the growth of plants, which inhibition is
PT antagonised by gibberellin, used to confer a dwarf phenotype
XX
PS Disclosure: Fig 7b; 88pp; English.
XX
CC The specification describes polypeptides encoded by the Rht gene (and
CC its homologues) that, when expressed in Triticum Aestivum, inhibit
CC growth of the plant. This growth inhibition is antagonised by
CC gibberellin. The products can be used to provide Rht expression in
CC plants, conferring a dwarf phenotype on a plant which is correctable
CC by treatment with gibberellin. In addition, the products can be
CC used to produce Rht mutant plants which are dwarfed compared with
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants

```

```

CC may be made by knocking out Rht or the relevant homologous gene in
CC the plant of interest. Plants may be made which are resistant to
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
CC dwarf but let crop plants grow tall. The present sequence is encoded by
CC the wheat Rht clone C15-1 cDNA sequence.
XX
SQ Sequence 425 AA:

Query Match 68.5%; Score 2189; DB 20; Length 425;
Best Local Similarity 100.0%; Pred. No. 7.3e-165;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 ARSSVEAAPVPAANAATPALPVPVVDVQEGIRLVHALLACAEVQDENSAAALVK 258
Db 1 arssvveaapvpaaanaatpalpvpvvdvqegirlvhalldacaeavqgenfsaaealvk 60
QY 259 QIPLLAASQSGAMRKVAAVFEGLARVRFRFRPODSSLLDAAFADLHAHFYESCPIYK 318
Db 61 qipllaasqsgamrkvaayfgealarrvfrfrpposslldaafadlhaHFYescpylk 120
QY 319 FAHFTANQAILAEAFACCRVHVVDGCIKOGMOWPALLQALARPGGPSPFRITGVGPPQ 378
Db 121 fahftangailaeafagccrvhvvdgikqgmwppallqalarrpgppspfriltvgppq 180
QY 379 DETDALQOYGMWLQAPAHRTIRVDFQYGLVAATLADLEPFMLQPGSEEDPNEPEVIAVN 438
Db 181 detdalqvgwklagfahltirvdfqyglvaatladdlpfmqpggedpnepevianv 240
QY 439 SVEFMHRLAQGALAEKVLGTVRAVPRRTVTVVEODPANHNSGFTLDRFTESLHYSTWFD 498
Db 241 svfemhrlaagpalekvlgtvrvavprvrtvvegeanhnsgftldrfeslhystmfd 300
QY 499 SLEGSSGGSPSEVSSGAAPAACTDOVMSEVYLGROICNVVACGEAERTERHETLQ 558
Db 301 slegssggspsevssgaapaaactdvmsevylygrqicnvacegaertelcrlq 360
QY 559 WRNRLGNAGFEVYHGSNMYKQASTLLAPAGDGYKVEKEGCTLTGWHTRPLIATSAM 618
Db 361 wrnrlgnagfevhlgsnaykqastllafagdykveekegcltlgwhtrpliatssaw 420
QY 619 RLGP 623
Db 421 rlgp 425

RESULT 4
AAE02560
ID AAE02560 standard; Protein: 587 AA.
XX
AC AAE02560;
XX
DT 10-AUG-2001 (first entry)
XX
DE A. thaliana transcription factor G308 homolog, G307.
XX
KW Plant transcription factor; phenotype; sugar sensing characteristic;
KW transgenic plant; plant yield; growth; germination; photosynthesis;
KW glyoxylate metabolism; respiration; pathogen response; wounding response;
KW cell cycle regulation; pigmentation; flowering; senescence; physiology;
KW storage organ; metabolism.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT Domain 323..339
FT /note="Conserved domain"
XX
PN W0200135725-A1.
XX
PD 25-MAY-2001.
XX

```





DR WPI: 2001-335999/35.  
 DR N-PSDB: AAD05776.  
 PT Nucleic acids encoding plant transcription factor polypeptides, useful  
 for altering the biochemical characteristics of plants e.g. corn,  
 potato and cotton plants -  
 PS Claim 4; Page 68-70; 127pp; English.  
 CC The present sequence is Arabidopsis thaliana transcription factor,  
 CC G307. The transcription factor is used for altering a plant's  
 CC biochemical characteristics. The transcription factor may be used to  
 CC alter the structure and developmental characteristics of plants such as  
 CC soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower,  
 CC alfalfa, sugar cane, turf, banana, blackberry, blueberry, strawberry,  
 CC raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant,  
 CC grapes, honey dew, lettuce, mango, melon, onion, papaya, peas, peppers,  
 CC pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon,  
 CC rosaceous fruits and/or vegetable brassicas. Transcription factors are  
 CC key controlling elements of biological pathways and altering expression  
 CC levels of 1 or more transcription factors can change entire biological  
 CC pathways in an organism. Therefore manipulating transcription factor  
 CC levels in plants offers great potential in agricultural biotechnology  
 CC for modifying a plant's traits. Transcription factor cDNA is useful in  
 CC gene therapy.  
 CC  
 XX  
 XX Sequence 587 AA:

Query Match 50.5%; Score 1612; DB 22; Length 587;  
 Best Local Similarity 53.7%; Pred. No. 4,7e-119;  
 Matches 340; Conservative 82; Mismatches 149; Indels 62; Gaps 13;

QY 1 MKREYODAGG---SGGGGGGSGSSEDKMMV--SAAAGEEVEDLLAAGYKVRASDNA 54  
 DB 1 mkrdhgfgrlsmhgtsessssiskkmvmvkkeedggmndellavlygkvrssema 60  
 QY 55 DYAKLEOLEMAMGAGVAGAAPDSFATHATDTVHNPPDISWVSMSLNAAPP 114  
 DB 61 evalleqletmm-----snvedgishlatdthvhpelyswldmnlsehn--pp 110  
 QY 115 PLPPAPQLMASTSYTSGGYFDLPSPVD--SSSIALPRIPBPAGATPA-DLSAD 170  
 DB 111 pjp-----aasnqldpvlpspeicgfpesdydkvlpnaalyqfpalssss 157  
 QY 171 SVRDKRMRTGSSSTS--SSSSSSSLGSGARSVVEAAPVAAANATPAAPVVDVQ 228  
 DB 158 smngqkrlkscspdsmyststgltqvgivltvtlttttaaeestr--vllvdsq 215  
 QY 223 ENGILVHALLACAAVQOENLSAABALVKQIPLAASGGAMRKVAAYFGEALARRVR 288  
 DB 216 engvvlvahlmcaaalqgnltlaealvkqigclavsqagamrkvaetyfaealarlyr 275  
 QY 289 FRPPDSSLDAAPDLHAHFESCPYLFKFAHTPANQALILEAFACGRVHVVDFFGIRKG 348  
 DB 276 l--spppqgldchldtclqmhyecopylkfahftangallaeefgkkrvvhidfdmng 333  
 QY 349 MOWPALQALALRPGSPFSRLTGVGPDPDETDALQGVGMKLQAFPAHTIRVDQYRGIV 408  
 DB 334 lqwpalmqalalreggpfrflltqgppadnsdhlhevgcklaqlaeahvefeyrgtv 393  
 QY 409 AATLADLEPFMLQPEGEENPEPEYIANVSYFEMHRLAOGCALEKVLGTVRAVPRIV 468  
 DB 394 anlsladldasml-----elrps--dteavavsvfelhklilgprgilekvigvkkqlprvlf 448  
 QY 469 TVVEEDANNSGTFIDRFESLHYSTFMDPSLEGSSGGSPSEVSSGAAPAAACTDOV 528  
 DB 449 tvvegesnmgldrfidrfeslhytstfdslseg-----vpsqdkv 490  
 QY 529 MSEVYLGROIQNVVAGEAERTERHETLQOMNRNLGNAGFETVHLGSNAVKAOSTLLALF 588  
 DB 491 msevylgkqiclnvacegdvrerhetlsqwgvrfgsgslapahlgnsafkqsmllsvf 550

QY 589 AGDDGYKEEGCLTGLMHTPLIATSAMRLA 621  
 DB 551 msqgyrveesngclmghwtrpilittsawkls 583

RESULT 6  
 AAM30792  
 ID AAM30792 standard; Protein; 532 AA.  
 XX  
 AC AAM30792;  
 XX  
 DT 19-MAR-1998 (first entry)  
 XX

DE Arabidopsis thaliana gibberellin insensitivity gai gene product.  
 XX  
 XX gibberellin insensitivity; gai; plant growth inhibition;  
 KW dwarf phenotype; lodging resistance; increased yield;  
 KW flowering regulation; bolting inhibition; spinach; lettuce;  
 KW antibody; identification; probe; primer; antisense; sense;  
 KW expression regulation; co-suppression; rice;  
 KW Bakane disease resistance.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 XX WO9729123-A2.  
 PD 14-AUG-1997.  
 XX  
 PF 12-FEB-1997; 97WO-GB00390.  
 XX  
 PR 12-FEB-1996; 96GB-0002796.  
 XX  
 PA (INNE-) INNES CENT INNOVATIONS LTD JOHN.  
 XX  
 DR Carol P, Harberd NP, Peng J, Richards DE;  
 XX N-PSDB: AAT91937.  
 XX

PT Nucleic acid encoding gibberellin inhibitor GAI and related  
 PT antisense sequences - used to create tall, or particularly, dwarf  
 PT plants, especially crops such as maize, rice and wheat  
 XX  
 PS Claim 1; Fig 4: 76pp; English.  
 XX

CC The present sequence is the Arabidopsis thaliana gibberellin  
 CC insensitivity (gai) gene product (GAI), the expression of which  
 CC inhibits plant growth. However the inhibition is antagonised by  
 CC gibberellin (GA), while gai expression confers a dwarf phenotype  
 CC that is insensitive to GA. Manipulating gai and GAI expression can  
 CC produce tall or dwarf plants, particularly the latter for increased  
 CC resistance to lodging and increased yield. It may also allow  
 CC regulation of flowering, i.e. plants remain in the vegetative state  
 CC until treated with GA, useful to inhibit bolting in spinach and  
 CC lettuce. GAI can be used to raise specific antibodies for  
 CC identifying homologous proteins or genes in other species. gai  
 CC fragments can also be used as probes or primers to identify and  
 CC clone related sequences, or in the preparation of antisense or  
 CC sense expression regulating (co-suppressing) sequences. Rice plants  
 CC that express GAI may be resistant to Bakane disease. Manipulation  
 CC of gai and GAI makes it possible to tailor the degree of dwarfism  
 CC and GA sensitivity to particular crops or situations.  
 XX

SQ Sequence 532 AA:

Query Match 49.9%; Score 1594; DB 18; Length 532;  
 Best Local Similarity 54.3%; Pred. No. 1.1e-117;  
 Matches 332; Conservative 71; Mismatches 106; Indels 102; Gaps 11;

QY 22 EDKMWVSAAGSEGEVEDLLAAGYKVRASDMADVAKLEOLEMAMGAGVAGAAPDS 81  
 DB 11 dktmmneedgngmndellavlygkvrssemdavqkleglevmm-----snvedd- 63



PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160817.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160880.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 49.9%; Score 1594; DB 21; Length 533;  
Best Local Similarity 54.3%; Pred. No. 1,1e-117;  
Matches 332; Conservative 71; Mismatches 106; Indels 102; Gaps 11;

QY 22 EDKMWSAAGGESEVDELIALAGYVRASDMADVOKLEOLMAMGMGVGAGAPDDOS 81  
Db 12 dkltmmneeddqngmdellavlygvkrssmadvaqkleglewm-----snvgdd- 64  
QY 82 FATHLATDVHVPNPTDLSWVESMSELNAPPPPLPAPQINASTSTVYSGGYFDLP 141  
Db 65 -lsqletcvhynpaelylvldsmldlpp----- 94  
QY 142 SVDSSSIYALRPISPPAGATAPADLSADSDPKRMRTGSGSTSSSSSSSLGGG--- 198  
Db 95 ---ssnaeydlkalp-----gdell-----ngfaldaasasnqg99gdc 130  
QY 199 -----ARSSVFEAPPAANAANFALPVYVVDPOEAGIRLVHALLACARVGOENT 250  
Db 131 ylttnkrlkcsngvve-----ttataestrhvlyvdsqengvrlvnaillacaaavqkenl 185  
QY 251 SAAEALVKQIPLLAASOGAMRKRYAIFGEALARVFRFRPODSSILDAFADLLHANF 310  
Db 186 tvaeealvkdqglfaveqdgjamrkvateyfaealarllyrl--spsgpidsldtqlgmhf 243  
QY 311 YESCPYIKFAHFTANQALILEAFAGCRHVYVDFGIRGOMQWPAALLQALIRPGPPSFKL 370  
Db 244 yetcpylikfahftaanaqalleafgkkrvhvldfmsqglqwpalmaqalalrpgppvfrl 303  
QY 371 TGWGPPOPETDALQOVGKMLAQFAHTIRVDPOYRGIVATLADLPFPMGPGEEDPNE 430  
Db 304 tglpppadnfdylinevgcklanlaeealneteyrgfvanctladlasml-----elrps- 338  
QY 431 EPEVIAVNSVFEEMHRLLAOPGALEKVLGTAVRAVPRIVTVBOEAMHNSGTFLDRFESTL 490  
Db 359 elesvaavnsvfelhkllgypgaldklygvnqklpelctfvegeshnnaplfdifdtesl 418  
QY 491 HYSTMPDSLGGSGGSGGSEVSSGAAAPAAAGTQVWSEVYVLGKQICNVVACEGAERT 550  
Db 419 hystclfdslsleg-----vpsgdkvmseylygkqlcnvacadpdrv 460  
QY 551 ERHETLGOVNRNLGNAGFETVHILGSNAYKQASLTLAFAGDGSKYVEKRGCTLGMHNR 610  
Db 461 etnelisqwnrtfqsagfaaahignafkqasmllalfnnggyrveesdgclmlywhtr 520  
QY 611 PLIATSAWRLA 621  
Db 521 pliatstawkls 531

RESULT 8  
AAE02545  
ID AAE02545 standard; Protein: 532 AA.  
XX  
AC AAE02545;  
XX  
D7 10-AUG-2001 (first entry)  
XX  
DE A. thaliana transcription factor G308.  
XX  
KW Plant transcription factor; phenotype; sugar sensing characteristic;  
KW transgenic plant; plant yield; growth; germination; photosynthesis;  
KW glyoxylate metabolism; respiration; pathogen response; wounding response;  
KW cell cycle regulation; pigmentation; flowering; senescence; physiology;  
KW storage organ; metabolism.  
XX  
OS Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers  
FT Domain 270..274  
FT /note="Conserved domain"  
XX  
PD MO200135725-A1.  
XX  
PD 25-MAY-2001.  
XX  
PF 14-NOV-2000; 2000MO-US31414.  
XX  
PR 17-NOV-1999; 99US-0166228.  
PR 17-APR-2000; 2000US-0197899.  
PR 22-AUG-2000; 2000US-0227439.  
XX  
PA (MEND-) MENDEL BIOTECHNOLOGY INC.  
PA (JIAN/) JIANG C.  
PA (HEAR/) HEARD J.  
PA (PINE/) PINEDA O.  
PA (PILG/) PILGRIM M.  
PA (ADAM/) ADAM L.  
PA (RIEC/) RIECHMANN J L.  
PA (YUGG/) YU G.  
PA (SAMA/) SAMAH R.  
PI Jiang C, Heard J, Pineda O, Pilgrim M, Adam L, Riechmann JL;  
PI Yu G, Samaha R;  
DR WPI: 2001-335977/35.  
DR N-PSDB: AAD06646.  
XX  
PT Nucleic acids encoding plant transcription factor polypeptides, useful  
PT for altering the sugar sensing characteristics of plants and increasing  
PT yield, e.g. corn, potato and cotton plants -  
XX  
PS Claim 4; Page 74-76; 151pp; English.  
XX  
CC The patent relates to polynucleotides encoding 35 plant transcription  
CC factors which may be used to modify phenotype associated with a plant's  
CC sugar sensing characteristics and increasing yield when their expression  
CC level is altered. Sugars are central regulatory molecules that control  
CC aspects of physiology, metabolism and development. Therefore the cDNAs  
CC and proteins of the invention are useful for modifying the growth and  
CC germination rates of plants, photosynthesis, glyoxylate metabolism,  
CC respiration, starch and sucrose synthesis and degradation, pathogen  
CC response, wounding response, cell cycle regulation, pigmentation,  
CC flowering and senescence of plants and for modifying sink-source  
CC relationships in seeds, tubers, roots, and other storage organs leading  
CC to an increase in yield. The transcription factor polynucleotides and  
CC polypeptides may be used to alter the structure and developmental  
CC characteristics of plants such as soybean, wheat, corn, potato, cotton,  
CC rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,  
CC blueberry, strawberry, raspberry, cantaloupe, carrot, calliflower,  
CC coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,

CC onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,  
CC tobacco, tomato, watermelon, rosaceous fruits and/or vegetable brassicas.  
CC The present sequence is an Arabidopsis thaliana transcription factor.  
XX  
SQ Sequence 532 AA:  
Query Match 49.8%; Score 1590; DB 22; Length 532;  
Best Local Similarity 54.2%; Pred. No. 2,2e-117;  
Matches 331; Conservative 72; Mismatches 106; Indels 102; Gaps 11:  
OY 22 EDKMWVSAAGBGEVEDELLAALGKVRASDMADVAQKLEULEMAMGCGAGAAPDDS 81  
DB 11 dkktmmmeedgngmdellavlygkvrsemadvaqlleglevmm-----snvgedd- 63  
OY 82 FATHLATDTVHNPRDLSWVESMLSELNAPPPLPPAPQALNASSVTGSGGFDDLP 141  
DB 64 -lsqlatetvhympaeltywldsmldlmp----- 93  
OY 142 SVDSSSIYALRPPSPAGATPAPADLSADSVDRPKMRGTSGSSSSSSSLGCG--- 198  
DB 94 ---ssnaeydkaip-----gdall-----ngfaidaassngsggddt 129  
OY 199 -----ARSSVEADPPVAAAANATPALPVVVDTQBPAGIRLVHALLACAEAVQENL 250  
DB 130 ytnkrlkcsngve-----ttataestrhvlyvdsqengyrlvhallicaeavqkenl 184  
OY 251 SAAEALVQIPLLAASOGGAMKVAAYFGEALARVFRFRPQDDSLUDAAFDLHMHF 310  
DB 185 lvaeealvqxqigflavsqigamqvafyfaealarlyrl--spsgspldhsldclgmif 242  
OY 311 YESCPYLFKFAHTANQALILEAFAGCRHVHVDFFGKQGMOPALQALALRPGPPSRL 370  
DB 243 yetcpylkfahftanqalileafagkkrvhvdfmsqglqwpalqalalrpgppvfrl 302  
OY 371 TGVGPPQPDETDALQOVGKRLAQFPHHTIRVDFQIRGLVAATADLEPPMLDPEGEDPRE 430  
DB 303 tgisppapdnfdylyevgcklahlaeahvefeyrgfvantladdaaml----elrps- 357  
OY 431 EPEVAVNSVFEHRLILQPGALEKLVTRAVRPRITVYVVEQANHNHSGRPLDFTSTL 490  
DB 358 elesvavnsvfeihklllrgpaiddkvilgvnqikpeltfveqeshnspilfdifteei 417  
OY 491 HYSTMPDLEGGSSGGSPSEVSSGAAAAAPAAAGTDQVMSVYLLGRLQICNVVACGAERT 550  
DB 418 hystclflsleg-----ypsgqdkmseyylgkqlcnvvaacdgdtrv 459  
OY 551 ERHETLGGMRRLGNAGFETVHLGSNATKQASTLALFAGSGGYVEKEGGLTGHTTR 610  
DB 460 erhetlsgwrrfsgagfaaahlgshatkgasmllalngesgyveesdgclmlywhtr 519  
OY 611 PLIATSAWRLA 621  
DB 520 pliatcawklis 530  
RESULT 9  
AAE01907  
ID AAE01907 standard; Protein: 532 AA.  
XX  
AC AAE01907;  
XX  
D7 31-JUL-2001 (first entry)  
XX  
DE Arabidopsis thaliana transcription factor, G308.  
XX  
KW Transcription factor; biochemical characteristic; controlling element;  
KW structural characteristic; developmental characteristic; gene therapy;  
KW agricultural biotechnology; plant trait modification.  
XX  
OS Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers

```

FT Domain                                270..274
ET /label= Conserved_domain
PN WO200136597-A1.
XX
PD 25-MAY-2001.
XX
PF 14-NOV-2000; 2000WO-US31344.
XX
PR 17-NOV-1999;    99US-O166228.
PR 17-APR-2000; 2000US-O197899.
PR 22-AUG-2000; 2000US-O227439.
XX
PA (MEND-) MENDEL BIOTECHNOLOGY INC.
PA (CREE/) CREELMAN R.
PA (YUGG/) YU G.
PA (ADAM/) ADAM L.
PA (RIEC/) RIECHMANN J L.
PA (HEAR/) HEARD J.
PA (SAMA/) SAMAH R.
PA (PIIG/) PILGRIM M.
PA (PINE/) PINEDA O.
PA (JTAN/) JIANG C.
XX
XX Creelman R, Yu G, Adam L, Riechmann JL, Heard J, Samaha R;
PI Pilgrim M, Pineda O, Jiang C;
DR WPI: 2001-33599/35.
NR N-PSDB: AAD05791.
XX
PT Nucleic acids encoding plant transcription factor polypeptides, useful
PT for altering the biochemical characteristics of plants e.g. corn,
PT potato and cotton plants -
XX
PS Claim 4; Page 114-115; 127pp; English.
CC The present sequence is Arabidopsis thaliana transcription factor,
CC G308, a homologue of G307. The transcription factor is used for altering
CC a plant's biochemical characteristics. The transcription factor may be
CC used to alter the structure and developmental characteristics of plants
CC such as soybean, wheat, corn, potato, cotton, rice, oilseed rape,
CC sunflower, alfalfa, sugar cane, turf, banana, blackberry, blueberry,
CC strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber,
CC eggplant, grapes, honey dew, lettuce, mango, melon, onion, papaya, peas,
CC peppers, pineapple, spinach, squash, sweet corn, tobacco, tomato,
CC watermelon, roseaceous fruits and/or vegetable brassicas. Transcription
CC factors are key controlling elements of biological pathways and altering
CC expression levels of 1 or more transcription factors can change entire
CC biological pathways in an organism. Therefore manipulating transcripion
CC factor levels in plants offers great potential in agricultural
CC biotechnology for modifying a plant's traits. Transcription factor cDNA
CC is useful in gene therapy.
XX
SQ Sequence      532 AA:
   Query Match              49.8%; Score 1590; DB 22; Length 532:
Best Local Similarity     54.2%; Pred No. 2.2e+117;
Matches 331; Conservative                               106; Indels 102; Gaps 11;
   22 EDKMYVAAGGEFEVDLIALGYRVASDADVAAOKLEQLEMAMGMGVGAAPAADD 81
       :|||::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  ||
Db 11 dkktmmmeeddngmdellavlygkxvrssemadraqrkeglevm-----snvgdedd- 63
       :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
   82 FATHLATDTVRHNPDDLSSVESMTSELNAPPPLPPAPQLNASTSVTGSGGYFDLP 141
       :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 64 -tsqlatetlvhyhpaeljtwldsmiltndmp-----                      93
   142 SVDSSTSYALPRISPGATAPADLSADSVDPKRMRTSGSSSTSSSSSSSIgg--- 198
       :||::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 94 ---ssnaegdlaiip-----gdail-----nqfalddsassnpgggagdc 129
       :||::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
   199 -----ARSSVEAAPPVAAAANAATPALPVVVVDTOENGIRLVHALACAEAVQEENTL 250

```

Accession	Gene	Protein	Length (aa)	Source
D6	130 yltcnrk1kcsngvve	-----tttaetstrhvtlvdsgengvtrivthalaceaevgkenl	184	
Oy	251 SAAEELVQIOLLLAASOGAMRKAAAYFGAELARVRFRFPQPPSSLLDAAFDLHAHF	310		
D6	185 tvaeealvqgiflavsqdgmqrqvatyfaealarrryrl--spsqpsldhsldtlgmhf	242		
Oy	311 YESPCYLFKAHFANQALILEAFACRCRRVHVDFGICQGMOMPALQALALRPGSPSEFRL	370		
D6	243 yetccp1kfahftangalleafgqkkrvhvdfmsqglqwpalmqalalrpgspvfvrl	302		
Oy	371 TGVGPPQPDENDDALQOVVGMKLAQFAHTRINDFOYRGVLVAATLADLEPFMIQPEGEEDPNE	430		
D6	303 tgi9ppapdnfdy1hevygcklahlaealhvefeygrfvantladiadsmtl---elrps-	357		
Oy	431 EPEVLAANVSVEEMHRLALQOPALEKVLGTVAAPRPRIVTVEQDANNSGTFIDRFESTL	490		
D6	358 e1esvaavsvfvelhkhllgrpaikdvlgvvnqikpelfltvegesnmspfidrfestl	417		
Oy	491 HYSSTMFDSLGGSSGGGSPSEVSSGAAAAAPAAAGTDQVMSEVYLGRQICNVVACEGART	550		
D6	418 hystllfslsleg-----vpsgqdkvmsevylygkq1cnvvaedgpdv	459		
Oy	551 ERHEFLGGMKRNRLGNAGREYTHLGSNATKQASTLLALFAGGDGKVKVEKRGCLTLGNHTR	610		
D6	460 erhetlsgvtrnrffsgafaaahgsnaikqasmllafnnggeyrvveesdgclmlyghtr	519		
Oy	611 PLIATSAWRLA 621			
D6	520 pliaatsawkls 530			
RESULT 10				
AC	AGG38577			
XX	ID AGG38577 standard; Protein: 517 AA.			
XX	AGG38577;			
DT	18-OCR-2000 (first entry)			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 47612.			
XX				
KM	Protein identification; signal transduction pathway; metabolic pathway;			
KW	hybridisation assay; genetic mapping; gene expression control; promoter;			
XX	termination sequence.			
OS	Arabidopsis thaliana.			
XX				
PN	EP1033405-A2.			
XX				
PD	06-SEP-2000.			
XX				
PF	25-FEB-2000; 2000EP-0301439.			
XX				
PR	25-FEB-1999; 99US-0121825.			
PR	05-MAR-1999; 99US-0123180.			
PR	09-MAR-1999; 99US-0123548.			
PR	23-MAR-1999; 99US-0125788.			
PR	23-MAR-1999; 99US-0125788.			
PR	29-MAR-1999; 99US-0126765.			
PR	01-APR-1999; 99US-0127462.			
PR	06-APR-1999; 99US-0128234.			
PR	08-APR-1999; 99US-0128714.			
PR	16-APR-1999; 99US-0129845.			
PR	19-APR-1999; 99US-0130077.			
PR	21-APR-1999; 99US-0130449.			
PR	23-APR-1999; 99US-0130510.			
PR	23-APR-1999; 99US-0130891.			
PR	28-APR-1999; 99US-0131449.			
PR	30-APR-1999; 99US-0132048.			
PR	30-APR-1999; 99US-0132407.			
PR	04-MAY-1999; 99US-0132484.			
PR	05-MAY-1999; 99US-0132485.			

PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144336.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147312.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147316.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149729.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 27-AUG-1999; 99US-0151303.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157755.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.

PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161922.
PR	28-OCT-1999;	99US-0161933.
PR	29-OCT-1999;	99US-0162142.

Query Match	49.7%;	Score 1589;	DB 21;	Length 517;
Best Local Similarity	55.0%;	Pred. No. 2.6e-117;		
Matches 330;	Conservative 69;	Mismatches 99;	Indels 102;	Gaps 11;

QY	33	EGEEVDLEALAGYKVRASMDADVNOQKLEOLEMAAGMGVCGAARPDSEFATHLATDYN	92
Db	7	dngngdelldavlygvrsvssemadvqklegleavm-----snvgdd--lsqleteth	58
QY	93	YNPTLSSWVSMSELNAAPPELPPALQNLASTSSVTGSGGFEDLPSSVDSSTLYL	152
Db	59	ynpaelylwldsmldtlrpp-----snaeydl	86
QY	153	RPISPAQATAPADLSDSVDRDPKRMRTGSSSTSSSSSSSLGGC-----ARS	201
Db	87	kaiP-----gdaIl-----nqfaiassassnqggsgdtytnkrllksn	125
QY	202	SVVEAPRVAAANATRALPVVYVDTQEGATLTVALLACAAVOENLSAEALYKQIP	261
Db	126	gvve-----ttataestcrhvvldvsgengvrlvhalIacaaevqkealIvkaIvKqIq	180
QY	262	LLAASOGGAMRKVAAYFGEALARVYFRFRPOPDSLLDAFADLLHAHFYECSPYIKFPAH	321
Db	181	flvsgqIasamtkvatyIfacealarrlyrl--spsgspidhsIsdltqmhIyetcpylkfiah	238
QY	322	FTANALILEAFAGCRHVHVDFGIKQGMQWPAALLQALALRPGPSFRLLTGVGPPODET	381
Db	239	ftanqallleaIeqgkkrvhvldfsmsgqIqwpalmqalalrpgppvfrltIqygpaaDnf	298
QY	382	DALQOVGKMLQAFATTINVDPOYRBUVAATLADLEPFLMQIPGEGEDPNEPEVIYAVNSVF	441
Db	299	dylvevgkllhlaealIveIeyIgfvancladIasml---elrps--elasesvansyf	353
QY	442	EMHRLIAPGALIEKVLGTVRAVRPRITVYVEQEAHNHNSGTFIDRFTESLHYSTMPDSIE	501
Db	354	elhklIgrpgaldkrlgyrvnqdkpelIcfevgeshmspIdIartfesiInhyStIIdsiE	413
QY	502	GGSGGGPGSEVSSGAAAPAAAGTDQVNSEVYILGRQICNVVACEGAERTERHETTLQWNN	561
Db	414	g-----vpsgqdkvmseyIlgxqIcvvacadpdrverheltIsqrn	455
QY	562	RLGNAGFETVLGSMAYKQASLTLLAFNGDQCYKYNKEKGCITLGMHRRPLATSMKRLA	621
Db	456	rfgsegIfaaahIgsnaIqfxasmlalIinggeyIrvkesdgclmlIghtrplIatsawklS	515
RESULT	11		
AAAG38576	ID	AAAG38576 standard; Protein: 518 AA.	
AAAG38576;			
XX	18-OCR-2000	(first entry)	
XX	Arabiopsis thaliana	protein fragment SEQ ID NO: 47611.	
XX	Protein identification; signal transduction pathway; metabolic pathway;		
XX	hybridisation assay; genetic mapping; gene expression control; promoter;		

KM	termination sequence.
XX	
OS	<i>Arabidopsis thaliana</i> .
XX	EP1033405-A2.
PN	
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439
XX	
PR	25-FEB-1999; 99US-0121825
PR	05-MAR-1999; 99US-0123348
PR	03-MAR-1999; 99US-0125318
PR	23-MAR-1999; 99US-0125748
PR	25-MAR-1999; 99US-0126564
PR	29-MAR-1999; 99US-0126785
PR	01-APR-1999; 99US-0127462
PR	06-APR-1999; 99US-0128234
PR	08-APR-1999; 99US-0128714
PR	16-APR-1999; 99US-0129845
PR	19-APR-1999; 99US-0130077
PR	21-APR-1999; 99US-0130049
PR	23-APR-1999; 99US-0130510
PR	23-APR-1999; 99US-0130891
PR	28-APR-1999; 99US-0131449
PR	30-APR-1999; 99US-0132048
PR	30-APR-1999; 99US-0132407
PR	04-MAY-1999; 99US-0132484
PR	05-MAY-1999; 99US-0132485
PR	06-MAY-1999; 99US-0132486
PR	07-MAY-1999; 99US-0132487
PR	07-MAY-1999; 99US-0132863
PR	11-MAY-1999; 99US-0134256
PR	14-MAY-1999; 99US-0134218
PR	14-MAY-1999; 99US-0134219
PR	14-MAY-1999; 99US-0134321
PR	14-MAY-1999; 99US-0134370
PR	18-MAY-1999; 99US-0134768
PR	19-MAY-1999; 99US-0134841
PR	20-MAY-1999; 99US-0135124
PR	21-MAY-1999; 99US-0135529
PR	24-MAY-1999; 99US-0135623
PR	25-MAY-1999; 99US-0136021
PR	27-MAY-1999; 99US-0136392
PR	28-MAY-1999; 99US-0136782
PR	01-JUN-1999; 99US-0137222
PR	03-JUN-1999; 99US-0137528
PR	04-JUN-1999; 99US-0137702
PR	07-JUN-1999; 99US-0137724
PR	08-JUN-1999; 99US-0138040
PR	10-JUN-1999; 99US-0138649
PR	10-JUN-1999; 99US-0138847
PR	14-JUN-1999; 99US-0139119
PR	16-JUN-1999; 99US-0139452
PR	16-JUN-1999; 99US-0139453
PR	17-JUN-1999; 99US-0139492
PR	18-JUN-1999; 99US-0139454
PR	18-JUN-1999; 99US-0139455
PR	18-JUN-1999; 99US-0139456
PR	18-JUN-1999; 99US-0139457
PR	18-JUN-1999; 99US-0139458
PR	18-JUN-1999; 99US-0139459
PR	18-JUN-1999; 99US-0139460
PR	18-JUN-1999; 99US-0139461
PR	18-JUN-1999; 99US-0139462
PR	18-JUN-1999; 99US-0139463
PR	18-JUN-1999; 99US-0139750
PR	18-JUN-1999; 99US-0139763
PR	21-JUN-1999; 99US-0139817
PR	22-JUN-1999; 99US-0139899
PR	23-JUN-1999; 99US-0140353
PR	23-JUN-1999; 99US-0140354
PR	24-JUN-1999; 99US-0140695









```

DB 192 -----trpllvdsqdgvrivthalmacaevgsnltlaaalkqgflvsgagam 244
QY 272 RKYAAVFGAALARVRFRPDPSSLLDAFADLHAHFESCPYKFAHTANQAILEA 331
DB 245 rkvalyfaealarriyrl--sppqtqidhsldtlgmhfetecpylkfahftangallea 302
QY 332 FACGRVHVWDFCIKQGMQMPALLQALARPGRPSPFRITGVPPPPDEPDALQOVGMKL 391
DB 303 fegkkrvhwidismngqlqwpalmgaalreggsprfritlgppaadnsdhlhevackl 362
QY 392 ADFAHTRVDFYRGVLAATLADLEPFMLQPEGEEDPNEPEEYVIANSVYEMHRLAQPg 451
DB 363 aqlaeelhefeyrgfvansladidasml-----elrps-eteavaavnsvelhkllyrtg 417
QY 452 ALEKVLGTVRAVRPRIVYVEQEAHNSGTFIDRFTESLHYVSTMPDSELGSSGGGPPSE 511
DB 418 gtekvrgvkkqlpvrflvtvegesnhngpvlfdrtfeslhyastlidsleg----- 468
QY 512 VSSGAAAPAAAGCTDQVMSFVYLGROIQNVVACGAEERPERHETLGOMWRRLGNAGFEV 571
DB 469 -----apssqdkvmeevylgkqicnlvacegpdvrerheltlsgwsmrfgssgfapa 519
QY 572 HLGSNAYKQASTLLALFAGGDDGYKVEKEGCLTIGWHTRPLIATSAMRLA 621
DB 520 hlgsmafkqastllalfnggegyrvekngcmlmswhtprlittsawklis 569

RESULT 15
AAB28574
ID AAB28574 standard; Protein; 587 AA.
XX
AC AAB28574;
XX
DT 09-FEB-2001 (first entry)
XX
DE Arabidopsis SCLa8.
XX
KW Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;
KW Transgenic plant; cell division; molecular marker; herbicide resistance;
KW salt resistance; pathogen resistance; insect resistance.
XX
OS Arabidopsis thaliana.
XX
PN MO200053723-A2.
XX
PD 14-SEP-2000.
XX
PF 07-MAR-2000; 2000MO-US05875.
XX
PR 10-MAR-1999; 99US-0265585.
XX
PA (UYNX ) UNIV NEW YORK STATE.
PI Benfey PN, Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L,
PI Helariutta Y, Bruce W, Lim J;
XX
XX WPI; 2000-594315/56.
DR N-PSDB; AAC65291.
XX
PT Scarecrow gene useful for producing transgenic plants expressing genes
PT whose product increases starch, lignin or cellulose biosynthesis and
PT confers herbicide, pathogen or insect resistance
XX
PS Claim 14; Fig 13; 200pp; English.
XX
CC The present sequence is given in a specification relating to the
CC structure and function of a regulatory gene, SCARECROW (SCR).
CC SCARECROW-like (SCL) genes encoding proteins containing an amino acid
CC sequence similar to the sequence of MOTIF III (WHII) of Arabidopsis SCR
CC protein are disclosed. SCR and SCL genes are useful for producing
CC transgenic plants whose cell division is modified and root and/or stem
CC development and gravitropism of stem or hypocotyl is altered. Cell

```

CC division is increased or decreased in roots resulting in thicker or  
 CC thinner root development. The transgenic plants are useful for expressing  
 CC a gene of interest encoding a gene product that confers herbicide, salt,  
 CC pathogen or insect resistance in root or embryos and genes encoding  
 CC starch, lignin or cellulose biosynthesis in shoots. The SCR gene also  
 CC confers less susceptibility to lodging in the transgenic plants than a  
 CC wild-type plant. SCR gene sequences are also useful as molecular markers  
 CC for a quantitative trait e.g. root or gravitropism trait in molecular  
 CC breeding of crop plants.

XX  
 XX Sequence 587 AA;

Query Match 47.7%; Score 1522; DB 21; Length 587;  
 Best Local Similarity 52.0%; Pred. No. 6,1e-112;  
 Matches 329; Conservativity 77; Mismatches 165; Indels 62; Gaps 14;

```

QY 1 MKREYODAG-----SGGGGGMGSSSEDKMYV--SAAAGEEYDELLAALGYKVRASDMA 54
DB 1 mkrtdhngfggrlshngtssssssskdkmmvkkedgggmddellavlygykvrsema 60
QY 55 DVAQKLEOLEMAMGMGVGAGAAPDQSEFATHLATDYVYVNPDTLSSWESMLSELNAPP 114
DB 61 evalkklqetlmm-----snaqedgl-shlatdaahypseilyswldmlseln--pp 110
QY 115 PLRPPAQLMNSTSTVYSGGYFDLPPSYD---SSSITVLRPIPSFAGATARA-DLSAD 170
DB 111 plp-----assngldprlpspelcgfpxsdylkvipxnaiygfpaids 157
QY 171 SVBDPKRMRTGSSSTS--SSSSSSSLGGARSSVVEAPPAVAAANATPALPVVVDTO 228
DB 158 smngnkrlkscsspsdmvntstfgtgyvigtvtlttttcaaaests--vilydsq 215
QY 229 EAGIRLVHALLACAAVQOENLSAALVROIPLLAASOGAMKRYAAYGALARRVR 288
DB 216 engvrlvthalmacaeealqgnltlaealvkgclavsgagmrkvalyfaealarriyr 275
QY 289 FRPOPSSLLDAFADLHAHFESCPYKFAHTANQAILEAFAGCRVRHVDFGIKOG 348
DB 276 1--sppngqidhclstdtlgmhfetecpylkfahftangalleaifegkkrvhwidismng 333
QY 349 MOWPALQALALRPGGPPSPFRITGVPPPPDEPDALQOVGMKLAQFHTIRVDFYRGVLY 408
DB 334 lqwpalmgaalaleegpptrfritlgippaadnsdhlhevacklqaelaehveleygfv 393
QY 409 AATLADLEPFMLQPEGEEDPNEPEEYVIANSVYEMHRLAQPALAEKVLGTVRAVRPRIV 468
DB 394 ansladldaasm-----elrps-dteavaavnsvelhkllygrxgiekvlgvbkqdtgdfh 448
QY 469 TVVEQEAHNSGTFIDRFTESLHYVSTMPDSELGSSGGGPPSEVSSGAAAPAAAGTDQV 528
DB 449 xwxrqpnhngpgflldgxtslhysslcxldslegxpn-----sgkkl 490
QY 529 MSEVYLGROIQNVVACGAEARTERHETLGOMWRRLGNAGFEVHLGSNAYKQASTLLALF 588
DB 491 msxylygxqicnlvacegpdvrerheltlsgwsmrfgssglapahlgsmatfkqasmllsvf 550
QY 589 AGGDDGYKVEKEGCLTIGWHTRPLIATSAMRLA 621
DB 551 nsqgyrveesngcmlmswhtprlittsawklis 583

```

Search completed: January 22, 2002, 16:25:59  
 Job time: 139 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:24:19 ; Search time 22.72 Seconds  
(without alignments)  
12.876 Million cell updates/sec

Title: US-09-485-529-103  
Perfect score: 76  
Sequence: 1 LNAPPPPPAPQ 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA: \*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/prodata/2/1aa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/prodata/2/1aa/backfilest1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	72.4	446	1	US-07-952-800-4
2	54	71.1	3118	2	US-08-457-273B-8
3	53	69.7	726	4	US-09-126-980-2
4	53	69.7	726	4	US-09-126-980-2
5	52	68.4	15	2	US-08-769-745-31
6	52	68.4	281	2	US-08-810-453-2
7	52	68.4	281	3	US-08-815-190A-2
8	52	68.4	281	4	US-09-290-640-25
9	52	68.4	281	4	US-09-479-524-3
10	52	68.4	281	5	PCT-US95-00362-2
11	52	68.4	288	3	US-08-545-196B-21
12	52	68.4	288	4	US-09-028-327-3
13	52	68.4	311	4	US-09-179-558-66
14	52	68.4	311	4	US-09-179-558-66
15	51	67.1	32	3	US-08-545-196B-63
16	51	67.1	170	5	PCT-US95-16806A-6
17	51	67.1	278	3	US-08-545-196B-19
18	51	67.1	294	1	US-08-512-986-6
19	51	67.1	294	1	US-08-361-806A-6
20	51	67.1	294	3	US-08-545-196B-9
21	51	67.1	294	4	US-09-028-327-4
22	51	67.1	294	5	PCT-US95-16806A-21
23	51	67.1	445	2	US-08-900-148-2
24	50	65.8	15	4	US-08-602-999A-430
25	50	65.8	15	4	US-08-602-999A-431
26	50	65.8	18	4	US-08-602-999A-371
27	50	65.8	20	4	US-08-602-999A-368

28	50	65.8	22	4	US-08-602-999A-448	Sequence 448, App
29	50	65.8	408	2	US-08-926-258-2	Sequence 2, App1
30	50	65.8	408	2	US-09-120-053-2	Sequence 2, App1
31	50	65.8	408	4	US-09-198-212-2	Sequence 2, App1
32	50	65.8	525	4	US-09-041-886-28	Sequence 28, App1
33	50	65.8	525	4	US-08-764-870-7	Sequence 7, App1
34	50	65.8	525	4	US-08-980-115-7	Sequence 29, App1
35	50	65.8	530	4	US-09-041-886-29	Sequence 2, App1
36	50	65.8	533	1	US-07-952-800-2	Sequence 30, App1
37	50	65.8	552	4	US-09-041-886-30	Sequence 31, App1
38	50	65.8	589	4	US-09-041-886-31	Sequence 16, App1
39	50	65.8	3119	1	US-08-246-982A-16	Sequence 16, App1
40	50	65.8	3119	1	US-08-453-265-16	Sequence 6, App1
41	50	65.8	3144	1	US-08-246-982A-6	Sequence 6, App1
42	50	65.8	3144	1	US-08-453-265-6	Sequence 42, App1
43	50	65.8	3144	2	US-08-457-273B-42	Sequence 21, App1
44	50	65.8	3144	3	US-08-556-419-21	Sequence 15, App1
45	50	65.8	3144	4	US-09-041-886-15	

## ALIGNMENTS

RESULT 1  
US-07-952-800-4  
; Sequence 4, Application US/07952800  
; Patent No. 5403925  
; GENERAL INFORMATION:  
; APPLICANT: OZATO, KEIKO  
; TITLE OF INVENTION: A NEW MEMBER OF THE NUCLEAR HORMONE  
; NUMBER OF INVENTION: RECEPTOR SUPERFAMILY AND A CDNA CLONE THEREOF  
; TITLE OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND AND TOWNSEND  
; STREET: 1 MARKER PLAZA, STEWART TOWER, SUITE 2000  
; CITY: SAN FRANCISCO  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/952,800  
; FILING DATE: 19920928  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Kenneth A  
; REGISTRATION NUMBER: 31,677  
; REFERENCE/DOCKET NUMBER: 15280-21-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-543-9600  
; TELEFAX: 415-543-5043  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 446 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-952-800-4

Query Match 72.4%; Score 55; DB 1; Length 446;  
Best Local Similarity 72.7%; Pred. No. 10;

Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 NAPPPLPPAP 12  
DB 31 SAPPPLPPAP 41

```
RESULT 2
US-08-457-273B-8
; Sequence 8, Application US/08457273B
; Patent No. 5849995
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael
; APPLICANT: Lin, Biaoyang
; APPLICANT: Nasir, Jamal
; TITLE OF INVENTION: Mouse Model for Huntington's Disease and
; TITLE OF INVENTION: Related DNA Sequences
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESS: Virginia Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 5849995th Carolina
; COUNTRY: US
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,273B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 3477-85A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3118 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-457-273B-8

Query Match      71.1%; Score 54; DB 2; Length 3118;
Best Local Similarity 81.8%; Pred. No. 83;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 APPPLPAPQ 13
  ||||| 1111
  Db 29 APPPPPPPPQ 39

RESULT 3
US-09-126-980-2
; Sequence 2, Application US/09126980
; Patent No. 6270956
; GENERAL INFORMATION:
; APPLICANT: Jones, Katherine
; APPLICANT: Wei, Ping
; APPLICANT: Gardner, Mitchell
; APPLICANT: Fang, Shi-Min
; TITLE OF INVENTION: A TRANSCRIPTIONAL COACTIVATOR THAT
; TITLE OF INVENTION: INTERACTS WITH TAT PROTEIN AND REGULATES ITS
; TITLE OF INVENTION: BINDING TO TAR RNA, METHODS FOR MODULATING TAT
; TITLE OF INVENTION: TRANSCRIPTION, AND USES THEREFOR
; FILE REFERENCE: SALK2231
; CURRENT APPLICATION NUMBER: US/09/126,980
; CURRENT FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/069,341
; EARLIER FILING DATE: 1997-12-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 2
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-126-980-2
```

```
Query Match      69.7%; Score 53; DB 4; Length 726;
Best Local Similarity 80.0%; Pred. No. 27;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 4 PPPPLPAPQ 13
  ||||| 11
  Db 717 PPPPLPPLPK 726
```

```
RESULT 4
US-09-476-482-2
; Sequence 2, Application US/09476482
; Patent No. 6284456
; GENERAL INFORMATION:
; APPLICANT: Jones, Katherine A.
; APPLICANT: Wei, Ping
; APPLICANT: Gardner, Mitchell
; APPLICANT: Fang, Shi-Min
; TITLE OF INVENTION: A TRANSCRIPTIONAL COACTIVATOR THAT
; TITLE OF INVENTION: INTERACTS WITH TAT PROTEIN AND REGULATES ITS BINDING TO TAR
; TITLE OF INVENTION: RNA, METHODS FOR MODULATING TAT TRANSCRIPTION, AND USES
; FILE REFERENCE: SALK2230-2
; CURRENT APPLICATION NUMBER: US/09/476,482
; CURRENT FILING DATE: 1999-12-30
; EARLIER APPLICATION NUMBER: 09/126,980
; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-476-482-2

Query Match      69.7%; Score 53; DB 4; Length 726;
Best Local Similarity 80.0%; Pred. No. 27;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 4 PPPPLPAPQ 13
  ||||| 11
  Db 717 PPPPLPPLPK 726
```

```
RESULT 5
US-08-769-745-31
; Sequence 31, Application US/08769745
; Patent No. 5955259
; GENERAL INFORMATION:
; APPLICANT: Holmes, Todd C.
; APPLICANT: Levitan, Todd B.
; APPLICANT: Brandeis University
; TITLE OF INVENTION: Mechanism for the Regulation of Ion
; TITLE OF INVENTION: Channel Activity
; FILE REFERENCE: BRU96-02
; CURRENT APPLICATION NUMBER: US/08/769,745
; CURRENT FILING DATE: 1996-12-19
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapien
; US-08-769-745-31
```

Query Match 68.4%; Score 52; DB 2; Length 15;  
Best Local Similarity 88.9%; Pred. No. 0.94;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
| | | | | | | |  
DB 3 PPPPLPPPP 11

RESULT 6  
US-08-810-453-2  
; Sequence 2, Application US/08810453  
; Patent No. 5858990

; GENERAL INFORMATION:  
; APPLICANT: Walsh, Kenneth  
; TITLE OF INVENTION: FAS LIGAND COMPOSITIONS FOR TREATMENT OF  
; TITLE OF INVENTION: PROLIFERATIVE DISORDERS  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: Federal Reserve Plaza, 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/810,453  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: S1237/7004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)720-3500  
; TELEFAX: (617)720-2441  
; TELEX: 343  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 281 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-810-453-2

Query Match 68.4%; Score 52; DB 2; Length 281;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
| | | | | | | |  
DB 50 PPPPLPPPP 58

RESULT 7  
US-08-815-190A-2  
; Sequence 2, Application US/08815190A  
; Patent No. 6046310

; GENERAL INFORMATION:  
; APPLICANT: Queen, Cary L.  
; APPLICANT: Schneider, William P.  
; APPLICANT: Vasquez, Maximiliano  
; TITLE OF INVENTION: Fas Ligand Fusion Proteins and Their  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/815,190A  
; FILING DATE: 11-MAR-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/614,584  
; FILING DATE: 13-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Apple, Randolph T.  
; REGISTRATION NUMBER: 36,429  
; REFERENCE/DOCKET NUMBER: 011823-006710US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 281 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-815-190A-2

Query Match 68.4%; Score 52; DB 3; Length 281;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
| | | | | | | |  
DB 50 PPPPLPPPP 58

RESULT 8  
US-09-290-640-25  
; Sequence 25, Application US/09290640  
; Patent No. 6204055  
; GENERAL INFORMATION:  
; APPLICANT: Dean, Nicholas M.  
; APPLICANT: Marcusson, Eric G.  
; TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling  
; FILE REFERENCE: ISPH-0351  
; CURRENT APPLICATION NUMBER: US/09/290,640  
; CURRENT FILING DATE: 1999-04-12  
; NUMBER OF SEQ ID NOS: 85  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 25  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-290-640-25

Query Match 68.4%; Score 52; DB 4; Length 281;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
| | | | | | | |  
DB 50 PPPPLPPPP 58

RESULT 9

US-09-479-524-3  
; Sequence 3, Application US/09479524  
; Patent No. 6268350  
; GENERAL INFORMATION:  
; APPLICANT: Barbara Guillem, Emilio  
; TITLE OF INVENTION: Polynucleotides For Inhibiting Metastasis And Tumor Cell Growth  
; FILE REFERENCE: B-37  
; CURRENT APPLICATION NUMBER: US/09/479,524  
; CURRENT FILING DATE: 2000-01-07  
; EARLIER APPLICATION NUMBER: US 09/170,948  
; EARLIER FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: US 60/062,733  
; EARLIER FILING DATE: 1997-10-23  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Wordperfect  
; SEQ ID NO: 3  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-479-524-3

Query Match 68.4%; Score 52; DB 4; Length 281;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
| | | | | | | |  
Db 50 PPPPLPPPP 58

RESULT 10  
PCT-US95-00362-2  
; Sequence 2, Application PC/TUS9500362  
; GENERAL INFORMATION:  
; APPLICANT: IMMUNEX CORPORATION  
; TITLE OF INVENTION: Ligand That Binds Fas Antigen  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: US  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple 7.1  
; SOFTWARE: Microsoft Word, Version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/00362  
; FILING DATE: 06-JAN-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/179,138  
; FILING DATE: 07-JAN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/190,559  
; FILING DATE: 01-FEB-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Anderson, Kathryn A.  
; REGISTRATION NUMBER: 32,172  
; REFERENCE/DOCKET NUMBER: 2805-WO  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 281 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US95-00362-2

Query Match 68.4%; Score 52; DB 5; Length 281;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
| | | | | | | |  
Db 50 PPPPLPPPP 58

RESULT 11  
US-08-545-196B-21  
; Sequence 21, Application US/08545196B  
; Patent No. 6080577  
; GENERAL INFORMATION:  
; APPLICANT: MELKI, JUDITH  
; APPLICANT: MUNNICH, ARNOLD  
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE  
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/545,196B  
; FILING DATE: 19-OCT-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FARACI, C. J.  
; REGISTRATION NUMBER: 32,350  
; REFERENCE/DOCKET NUMBER: 2121-110P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 288 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-545-196B-21

Query Match 68.4%; Score 52; DB 3; Length 288;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
| | | | | | | |  
Db 214 PPPPLPPPP 222

RESULT 12  
US-09-028-327-3  
; Sequence 37, Application US/09028327  
; Patent No. 6130064  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guegler, Karl J.  
; TITLE OF INVENTION: HUMAN SMN-LIKE PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.



CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/028,327  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0482 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-845-4166  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 288 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1857114  
US-09-028-327-3

Query Match 68.4%; Score 52; DB 4; Length 288;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
|||||||  
Db 214 PPPPLPPPP 222

RESULT 13  
US-09-179-558-66  
Sequence 66, Application US/09179558  
Patent No. 6180612  
GENERAL INFORMATION:  
APPLICANT: Hockensmith, Joel W.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
TARGETING DNA METABOLIC PROCESSES USING  
TITLE OF INVENTION: AMINOGLYCOSIDE DERIVATIVES  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/179,558  
FILING DATE: 27-OCT-1998  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: U.S. 09/060,470  
FILING DATE: 15-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 60/063,898  
FILING DATE: 31-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 9426-005-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-179-558-66

Query Match 68.4%; Score 52; DB 4; Length 311;  
Best Local Similarity 88.9%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
|||||||  
Db 12 PPPPLPPPP 20

RESULT 14  
US-07-945-283-2  
Sequence 2, Application US/07945283  
Patent No. 5352596  
GENERAL INFORMATION:  
APPLICANT: Cheung, Andrew K.  
TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants  
TITLE OF INVENTION: Involving The Ep0 and LfT Genes  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis P. Ribando  
STREET: 1815 No. 5352596th University Street  
CITY: Peoria  
STATE: IL  
COUNTRY: USA  
ZIP: 61604  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/945,283  
FILING DATE: 19920911  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ribando, Curtis P  
REGISTRATION NUMBER: 27976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 309-685-4011 ext. 513  
TELEFAX: 309-685-4128  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1958 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-945-283-2

Query Match 68.4%; Score 52; DB 1; Length 1958;  
 Best Local Similarity 88.9%; Pred. No. 91;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
 ||||| 1  
 Db 489 PPPPLPPP 497

## RESULT 15

US-08-545-196B-63  
 ; Sequence 63, Application US/08545196B  
 ; Patent No. 6080577  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MELKI, JUDITH  
 ; APPLICANT: MUNNICH, ARNOLD  
 ; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE  
 ; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY  
 ; NUMBER OF SEQUENCES: 65  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
 ; STREET: PO BOX 747  
 ; CITY: FALLS CHURCH  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22040-0747  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentln Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/545,196B  
 ; FILING DATE: 19-OCT-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: FARACI, C. J.  
 ; REGISTRATION NUMBER: 32,350  
 ; REFERENCE/DOCKET NUMBER: 2121-110P  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 205-8000  
 ; TELEFAX: (703) 205-8050  
 ; INFORMATION FOR SEQ ID NO: 63:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 32 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-545-196B-63

Query Match 67.1%; Score 51; DB 3; Length 32;  
 Best Local Similarity 72.7%; Pred. No. 2.5;  
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 NAPPPLPPAP 12  
 | | | | |  
 Db 6 NGPPPPPPPPP 16

Search completed: January 22, 2002, 16:24:20  
 Job time: 40 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:23:40 ; Search time 26.73 Seconds  
(without alignments)  
1775.411 Million cell updates/sec

Title: US-09-485-529-7

Perfect score: 3194

Sequence: 1 MKREYQDAGSGGGGGGSGMS.....TLGHHTPLIATSNWRLGP 623

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1608	50.3	587	2	D84426 hypothetical prote
2	1594	49.9	533	2	H86282 protein F10B6.34 l
3	1511.5	47.3	511	2	G96688 hypothetical prote
4	1450.5	45.4	523	2	T51475 RGA-like protein -
5	604.5	18.9	653	2	T47581 SCARECROW1 - Arabi
6	600.5	18.8	653	2	T51244 SCARECROW protein
7	572.5	17.9	482	2	E96540 hypothetical prote
8	558	17.5	526	2	E96542 SCARECROW-like pro
9	554	17.3	413	2	G84462 SCARECROW-like pro
10	551	17.3	658	2	D96656 hypothetical prote
11	512	16.0	313	2	T51242 Scil protein (limp
12	507	15.6	306	2	T51236 SCARECROW-like pro
13	498	15.6	593	2	E86347 hypothetical prote
14	464	14.5	352	2	T51234 SCARECROW-like pro
15	445.5	13.9	1336	2	T02736 probable SCARECROW
16	444.5	13.9	542	2	T46142 SCARECROW-like 7 (
17	443	13.9	808	2	T51232 SCARECROW-like pro
18	433.5	13.6	531	2	T04722 SCARECROW-like prote
19	429.5	13.4	284	2	T51241 SCARECROW-like pro
20	415	13.0	375	2	C71441 probable SCARECROW
21	402.5	12.6	325	2	T51235 SCARECROW-like pro
22	399.5	12.5	718	2	T02531 probable SCARECROW
23	396	12.4	483	2	T10552 hypothetical prote
24	389.5	12.2	583	2	T45597 SCARECROW-like pro
25	384	12.0	573	2	T51239 SCARECROW-like pro
26	344	10.8	410	2	T45848 hypothetical prote
27	304	9.5	623	2	T47874 SCARECROW-like pro
28	298.5	9.3	640	2	B84887 probable SCARECROW
29	293	9.2	486	2	E85433 SCARECROW-like pro

30	288.5	9.0	558	2	T01343 hypothetical prote
31	285.5	8.9	378	2	T51237 SCARECROW-like pro
32	187.5	5.9	205	2	T51233 SCARECROW-like pro
33	179.5	5.6	261	2	T51243 Scil protein (limp
34	160	5.0	133	2	T51240 SCARECROW-like pro
35	144.5	4.5	1032	2	D83637 serine/threonine p
36	139.5	4.4	1428	2	T13926 probable protein p
37	138	4.3	443	1	T38239 transcription fact
38	134.5	4.2	1151	2	T18535 high molecular mas
39	134	4.2	523	2	T36677 probable secretory
40	133.5	4.2	1420	2	T37781 probable cytoskele
41	132.5	4.1	1487	2	T02850 hypothetical prote
42	130.5	4.1	275	2	T04480 acyl-CoA oxidase h
43	129	4.0	1041	2	T15521 hypothetical prote
44	127.5	4.0	1446	1	A45344 immediate-early pr
45	127	4.0	399	2	I49754 homeobox protein -

#### ALIGNMENTS

RESULT 1	D84426	hypothetical protein At2g01570 [imported] - Arabidopsis thaliana
C:Species:	Arabidopsis thaliana (mouse-ear cross)	
C:Date:	02-Feb-2001	#sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession:	D84426	
R:Lin:	X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Beilto, M.I.; Town, C.D.; Fujii, C.Y. M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.B.; Umayam, L.; Tallon, euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter Nature 402, 761-768, 1999	
A:Title:	Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.	
A:Reference number:	AB4420; M01D:20083487	
A:Accession:	D84426	
A:Status:	preliminary	
A:Molecule type:	DNA	
A:Residues:	1-587 <STO>	
A:Cross-references:	GB:AE002093; NID:g3785986; PIDN:AAC6733.1; GSPDB:GN00139	
C:geneids:	A:gene: At2g01570	
A:Map position:	2	
Query Match	50.3%; Score 1608; DB 2; Length 587;	
Best Local Similarity	53.6%; Pred. No. 3.1e-92;	
Matches 339; Conservative 82; Mismatches 150; Indels 62; Gaps 13;		
QY 1	MKREYQDAGS-----SGGGGCGMGSSDDKMW--SAAAGEEVEDLLAAGYVRASDMA 54	
DB 1	MKRDHHPQGRLSNHGTSSTSSISIKDKMMVKKEDDGGNMDELAVLGTVRSSEMA 60	
QY 55	DVAQKLEOLEMAMGKMGVGAAPDDSFATHLATDVHYVPTDLSVSWESMLSELNAPP 114	
DB 61	EVALKLEQLETFM-----SNVQEDGLSHLATDVHYVPSLEYSLMDLMEELN--PP 110	
QY 115	PLRPADQLNASTSYVTGSGGYFDLPPSYD---SSSIVALRPISPAGATAPAD-DLSAD 170	
DB 111	PLP-----ASSNGLDPLVLPSPICGFPASDYLKYLPGNAIYQFPALDISSSS 157	
QY 171	SVADPKRMRTGGSSST--SSSSSSSLGGARSSVYEAPVAAAANATATLVVVVDTG 228	
DB 158	SNQNQRRLKSCSSPDSWVSTGTGTGGVIGTVTTTTTTTAAAGSTBS--VIVDSQ 215	
QY 229	ENGIRLVHALLACAEVVOENLSAEEALVQIPLLAASOGAMRKVAAUYGEALARVFR 288	
DB 216	ENGVRLVHALLACAEALQNNLTLEALVQICLAVSQGAMRKVATYTAELARIVR 275	
QY 289	FRDQPSLLDAAFADLLAHAFYECSPYKFAHFTANQALILEAFAGCRVHVVDFTKQG 348	
DB 276	L--SPQNDICHLSDPLQNHFFETCPYLKFAHFTANQALILEAFEGKRVHVIVDSNQG 333	
QY 349	MORPALLQALALRGGPRPRLTGVGPORDPDIDALQVQWKLQAFHNTIRVDFQVRGLV 408	

[illegible][illegible]

Qy	371	TGVPBPQDEPDALDOXGMZKIAORAHNIIRDFOYRGUVAATLADLEPMQDPSGEDPNE	430
Dd	304	TGIGRPAPDNFDYJHEWGCKLAHLAEAIHVEEFKEFGVANTLADLDMSML----	ELRPS- 358
Qy	431	EPEYIANSVFEMHRLLAOPALEKVLGTAVARVPRIVTVVEOEAHNNSCTFLDRFTESL	490
Dd	359	EIESAAVNVSFELKHLLGRPCALDKVLGVUNQIQPEIFITVVEOSNNNSPIFLDRFTESJ,	418
Qy	491	HUYTMDPSLEGSSGGGGSPSYSSGAANAAGCTODVMSEVYLGRICNVACEGAERT	550
Dd	419	HUYSTLEFDSLESG-----VPSQDDVMSEVYLGIKCINWACDGDYV	460
Qy	551	ERHETLGWMRRRLNAGFEYVHLGSNAKYKAOSTALLAFAGDGGYKYEKKGCSTLGNMTR	610
Dd	461	ERHETLSQMRRRFPGSAGRPAAHGISNAFKQAQSMILLALFNGBEGRYVESDSCGLMGWNR	520
Qy	611	PLIATSAMRLA 621	
Dd	521	PLIATSAMRLS 531	

RESULT 3  
G96688  
hypothetical protein T27F4.10 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: G96688  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
C.A.: Li, J.H., Li, Y., Liu, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: G96688  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-511 <STO>  
A:Cross-references: GB:AE005173; NID:g10092507; PIDN:AA612907.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: T27F4.10  
A:Map position: 1

	Query Match	47.38;	Score 1511.5;	DB 2:	Length 511;
	Best Local Similarity	52.33;	Pred. No. 2.5e-86;		
	Matches 324; Conservative	65;	Mismatches 118;	Indels 113;	Gaps
Qy	1 MKREYDAGSGGGGGSSEDKMNVSAAGBECEYEDLLAALGYVRASDAADVAOAKL	60			
Db	1 MKRHHNRESSAGAGG--SSMTTVIEEPAAG---VDLLVVLGVKRVSSMDADVAAHKL	54			
Qy	61 EOLEAMMGCVGAGAAPDDSFATLATDTVHYNPPTDLSSWVESMLSELNAPEPLPPAP	120			
Db	55 EQLEMYLGDG-----ISNLDEYTHYRPDLSCGWESMLSDLD--PTRIOEKP	100			
Qy	121 QLNASTSTVTGSGGYTDLPPSYDSSSITAYLRPIPSPATATAADLSADVDPKMRKT	180			
Db	101 -----DSEYDLRAI--PGSAVYPRD--EHVTRRSKRTRL	130			
Qy	181 GGSSTSSSSSSSLGGGARSSVVEAAPPVAAAAANATPALPYVVYVTDTOEGIRLVHALLA	240			
Db	131 ESSELSTRS-----VVVLDSQEGVRLVHALLA	158			
Qy	241 CAEAVOQENLSAALVKQJLPBLAASOGCAEMRKYAATFGALARAVRYRRFPQDDSSLDA	300			
Db	159 CAEYVOQNMLEALVALKHVGLASSQAQGMARKVATVFAGLARRIYRIPRODVALL-S	216			

QY	301	AFALHAAHYESCPYKFKHAFHANOAILFAFGGCRHVYVDFGIQOGQWPAILOAL	360
		: : : : :	
Db	217	SFSLTLDIHYESCPYKFKHAFHANOAILFEVFETACKVHYDIDGLHNGQWPAILOAL	276
		: : : : :	
QY	361	RPGPSPSRITGVGPPQPDSDALOOGWKLAOFARTIRYVDFQYBGLVAATLADLEPFL	420
Db	277	RPNGPPQFRLTGIQSLT-----IQEYMKLGGOLASTIGVNEEFKSIALNNISDLKPEML	332
QY	421	QPEGEEDPNEEPEVIAVSYFEMHRLAOPGALAEKYLGTVAAPRPIYTVVEQEAHHNSG	480
		: : : : :	
Db	333	-----DIRGLESAVANSYFELHRLLAHNGSIDKLTSLTKIRPDIYTVVEQEAHHNCT	386
QY	481	TFIDRFETSLHYHYSTYMDLSLEGSSGGGSPSEVSSGAAAPAAAGTDQVNSEVYILGRIQCN	540
Db	387	VFIDRFETSLHYHYSLEDSLE-----GPPSQ-----DRVNSELFLGQIILN	427
QY	541	VVACEGAERTERHETTLGOWBNRLGNAGFETVLHGSNAYYQOASTLLALFAFGDGQYVEKE	600
		: : : : :	
Db	428	LVACEGEDRYERHETTLQOMNRRFLGJGFKPVSIGSNAYYQOASMLLALYAGADGQYVEENE	487
QY	601	GCCLTGLGHTHTRPLATISAMRL 620	
Db	488	GCCLLGMOTRPLATISAMRI 507	

	RESULT	4	
	T51475		RGA-like protein - Arabidopsis thaliana
	N:Alternate names:		protein K3M16_60
	C:Species:		Arabidopsis thaliana (mouse-ear cress)
	C>Date:	18-Aug-2000	#sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
	C:Accession:	T51475	
	R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; M		
	Submitted to the Protein Sequence Database,	August 2000	
	A:Reference number:	225394	
	A:Accession:	T51475	
	A>Status:	preliminary	
	A:Molecule type:	DNA	
	A:Residues:	1-523 <SAT>	
	A:Cross-references:	EMBL:AJ391150	
	A:Experimental source:	cultivar Columbia; BAC clone K3M16	
	A:Genetics:		
	A:Map position:	5	
	A>Note:	K3M16_60	
	Query Match	45.4%;	Score 1450.5; DB 2; Length 523;
	Best Local Similarity	49.3%;	Pred. No. 1.5e-82;
	Matches 313;	Conservative 76;	Mismatches 115; Indels 131; Gaps 13;
QY	1 MKREYDA-----GSGGGGGGGSSEDKMMVSAAGEEVEDELLALGY 46		
	: :		
DB	1 MKRSHQETSVEEAPSMWEKLENGCGGG-----DDNM-----DEFNAVIGY 42		
QY	47 KYRASDMADVAOKLEOLEEMAMGCVGAGAARDDSFATHLATDVHYNPTDLSWVESML 106		
	:                 : : : :		
DB	43 KYRSSMDADVAOKLEOLEEVL-----SNDLASSNMFNDIYHNPSDLISGAQSML 93		
QY	107 SELNAPPPLPPAPOLNASTSYVTGSGGYFDLPSPVSSSITYALRPISPAGATPAD 166		
	:		:
DB	94 SDLNYYP-----DLDP-----NRICDLRP-----TDDE 118		
QY	167 LSAQSVPRPKRMRTGGSTSSSSSSSLGGGARSSVEAARPVAAANATPALPYYVD 226		
	:		
DB	119 CCSSNSNKRIIRUGPCWDVTSSTRS-----VVL 149		
QY	227 TOEAGIRLVHALLAACAFAVOQENTSAEALVKQIPILTAASOGARKVYAAYFGAELARR 286		
	:		
DB	150 IETGVSRVQLAVLCAEAIVOLENLTLADALVRKGLLAASAQGANGKVATYFAELARI 209		
QY	287 FRFRQPQPSLLDAAFADLLAHFTYESCPYLKFHAHTFANOAILAEFACRCRVHVVDGIK 346		
	:		
DB	210 YRI--HPSAAAIDPSFEEDILQNFYDSPCYLKFAHTFANOALLEAVTTSRRVVHVIDLEIN 267		

Qy	347	QGMOMPALQALAIRPGGPPSFRITLGTGPPQRPQRETDALQOVGKLNQFATTIVVDQYR	406
Db	268	QGMOMPALMQALAIRPGGPPSFRITLGTG--NPSNRSEIOELGKMLQALQALIVERFNG	325
Qy	407	LVAATLADLIEPMLQRPGEEDPNEEPVIAVNSYFEGHRIALQAPGALKEKVLGTAVRVRP	466
Db	326	LTTERLSDLIEDPMF-----ETRESEPTLVVNSYFELHPVLSQPGSIEKLLATVYKAKPG	379
Qy	467	IYVYVEGEANNSGTFPLDRFLFESLHYSTAFDLSLEGSSSGGSGSEVSSGAAAAPAAAGD	526
Db	380	LVTYVEGEANNGVFLDRFNEALHYYSSLFDSLEDL-----VYIPS--QD	423
Qy	527	QVMSEVYLGRQICVNWCEGAEARTERRHETLGOMENRNLGNAGFETVHLGSNAYKQASTLLA	586
Db	424	RYMSVYILGRQILNVATNEGSDRIERHETLLQWKKRKNGSGNFDVYNLGSAPFQAASLLA	483
Qy	587	LFPAGDGYKEKEGECCLTGMHPRPLIATSAWMLA	621
Db	484	LSGGSDGYRVEENDGSIMLAQWOTRPLTAASAWMLA	518

	RESULT	5	
	T47581		
	SCARECROW1 - Arabidopsis thaliana		
	N:Alternate names: protein F24B22.180		
	C:Species: Arabidopsis thaliana (mouse-ear cress)		
	C:date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000		
	C:accession: T47581		
	R:Bloembergen, H.; Nemes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quecietier, F.; Salanoubat, M.		
	submitted to the Protein Sequence Database, January 2000		
	A:Reference number: Z23016		
	A:Accession: T47581		
	A>Status: preliminary		
	A:Molecule type: DNA		
	A:Residues: 1-653 <BIOL>		
	A:Cross-references: EMBL:AL132957		
	A:Experimental source: cultivar Columbia; BAC clone F24B22		
	C:Genetics:		
	A:Map position: 3		
	A:introns: 511/2		
	A>Note: F24B22.180		
	Query Match	18.9%	Score 604.5; DB 2; Length 653;
	Best Local Similarity	30.0%; Pred. No.5.3e-30;	
	Matches 179; Conservative 84; Mismatches 202; Indels 131; Gaps 18;		
OY	79 DDFSPATLATDTVHYNPTDLSSWESMLSEL-----NAAPPPLPPAPQLMA 124	: :   :	:
Dd	129 DDSSSSASASTV-----WDATIRDLIHSTSVSIPOLLQNVRLIFPCNPRLGA 179	: :   :	:
OY	125 -----STSYTVTGGSGGYFDLPSPVSUSSSIYALRTPSPACATAPADISAD 170	:  :	:  :
Dd	180 LLEYRLRSIMLLDPSSSS-----PESPQTFEPFLYOISNNPS----- 216	:	:
OY	171 SVRPDKRRRTGSSSTSSSSSSSLGGARSVEANPAAANAANPALPYVVDTQEA 230	: :   :	:
Dd	217 ----FQQQQQHQQQQOQHKKPPPPIQQDEKRNSTDARPOPEYTAN--VPANQTNTAEA 270	: :   :	:
OY	231 -----GIRLVHALLACAENAVODENTSAEALKVKQIPLLAASOGAMRKV 274	: :   :	:
Dd	271 LRERKEIKRKQCODEGHILTLTLCCAENASADNLBEANKLLLEISQLSTPYGTSMQRV 330	: :   :	:
OY	275 AAYGEALARKVF-----RRRPQDSSLDAAFRDLIHAIFYSCPLYKAHF 322	: :   :	:
Dd	331 AAYSEASMSARLLNSCLGIYAALPSRMWPOTHSLSKVMSA----QVNGISPLVFRSHF 385	: :   :	:
OY	323 TANQAIIIEAFPGCRRVHVVDFGIRKQGMOMPALQLALRPGGPSPFLTGVPPODET 382	: :   :	:
Dd	386 TANQAIOAEFAFKEDSVNHITIDIMOGLOWRELFIITILSRGGGRPHVHVLTLGLG----TSME 441	: :   :	:
OY	383 ALQGVGMVLAQFAHTIRVDFQYRGVLVAATLADLEPFMIQDEGEDPNNEDEVIYAVNSVFE 442	: :   :	:

```

Db      442 ALQATGKRLDFADKLDPFPCPL-AEKVGNLDERLVN-----RKREAVAVH--WL 491
QY      443 MHRLLAQPGALEKVLGVRAVRPRIVVVEOEANHNSTGLDRFTESLHYSTFDSLEG 502
      492 QHSLYDVTGSDAHTLMLQRLAPRVVTVVEDDLSH-AGSFLGRFVEAIIHYTSALFDSL-- 548
QY      503 GSSGGGPFSEVSSGMAAAPAAAGTDQVNSEVYLGROIQNVVACEGAEERTERHETLIGQWRNR 562
      549 GASVGESE-----ERHVEEQQLSKEIRNVLAAGSPRSAGEVK-FESNREK 594
QY      563 LGNNGFETHLGSNAVYQASTLLALFAGDGYKVEKEGCLTIGMHRPLIATSAW 618
      595 MQCCGFGISLAGNAAATQATLLGMFP-SDGYTLVDNGTLKLMKDSLITLTAASW 649

RESULT 6
T51244
SCARECROW protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 28-Jul-2000 #sequence_revision 28-Jul-2000 #text_change 28-Jul-2000
C:Accession: T51244
R:di Laurenzio, L.; Mysoczka-Dillier, J.; Malamy, J.E.; Pysh, L.; Helariutta, Y.; Freshour
submitted to the EMBL Data Library, July 1996
A:Reference number: Z23538
A:Accession: T51244
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-653 <DIL>
A:Cross-references: EMBL:U62798; PIDN:AA06318.1
C:Genetics:
A:Gene: SCARECROW1
A:Introns: 511/2

```

```

Query Match      18.8%; Score 600.5; DB 2; Length 653;
Best Local Similarity 29.9%; Pred. No. 9.3e-30;
Matches 178; Conservative 84; Mismatches 203; Indels 131; Gaps 18;

QY      79 DDFATHTLATDTVYNYNPTLDSVWESMSEL-----NAPPPPLPAPOLNA 124
      129 DQSSSSSSASPTV-----WYDAIIRDLIHSTSVSIPQLIQNVKDIITPCNPILGA 179
      125 -----STSYTVGSGYFDLPSPVSSSSSYVALRPIPSGATAPADLSAD 170
      180 LLEKRLSLMLDSSSS-----DPSQTFEPLVQISNNP----- 216
      171 SVRPRKMRRTGSGSTSSSSSSSLGGARSSVVEAPPVAAANATPRLPVVVDIOEA 230
      217 ---PQQQQQHQQQQQHKPPPPPIQQOERENSTDAPOPETVAT--VPAVQTNAAE 270
      231 -----GIRLVHALLACAEVQENLSAAELVYKQIPLLAASOGAMRKV 274
      271 LREKKEIKRKQKODEBGLHLTLTLQCAEAVSADNLEBANKLLEISOLSPITYTSQORV 330
      275 AAYFGEALARRVF-----RRPOPDSLLDAAFADLHAHFYESCPLYKFAHF 322
      331 AAYFSEAMSRLLNSCLGIYAALPSRMMPPQTHSLKMSAF-----QYFNGISPLVKFSHF 385
      323 TANQALILEAFACRRVNVDPGKQGMWPAALLQALALRPGGPPSFRITGYPPOPDDET 382
      386 TANQALIDEAEKEEDSVHIITDIDIMQGLWPGFLFIILASRPGPPHVRITGLG---TSM 441
      383 ALQGVKLLQAFHTIVDPOYRGIVAAATLADLEPMLQPGGEDPNEPEVIAVNSVFE 442
      442 ALQATGKRLDFADKLDPFPCPL-AEKVGNLDERLVN-----RKREAVAVH--WL 491
      443 MHRLLAQPGALEKVLGVRAVRPRIVVVEOEANHNSTGLDRFTESLHYSTFDSLEG 502
      492 QHSLYDVTGSDAHTLMLQRLAPRVVTVVEDDLSH-AGSFLGRFVEAIIHYTSALFDSL-- 548
      503 GSSGGGPFSEVSSGMAAAPAAAGTDQVNSEVYLGROIQNVVACEGAEERTERHETLIGQWRNR 562

```

```

Db      549 GASVGESE-----ERHVEEQQLSKEIRNVLAAGSPRSAGEVK-FESNREK 594
QY      563 LGNNGFETHLGSNAVYQASTLLALFAGDGYKVEKEGCLTIGMHRPLIATSAW 618
      595 MQCCGFGISLAGNAAATQATLLGMFP-SDGYTLVDNGTLKLMKDSLITLTAASW 649

RESULT 7
E96540
hypothetical protein FILF12.22 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E96540
R:Thielogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: E96540
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <STO>
A:Cross-references: GB:AE005173; NID:99454566; PIDN:AAF87889.1; GSPDB:GN00141
C:Genetics:
A:Gene: FILF12.22
A:Map position: 1

```

```

Query Match      17.9%; Score 572.5; DB 2; Length 482;
Best Local Similarity 28.7%; Pred. No. 3.4e-28;
Matches 147; Conservative 79; Mismatches 171; Indels 115; Gaps 10;

QY      181 GGSSTSSSSSSSLGGARSSVVEAAPVAAANATPRLPVVVDIOEAGIRLVHALLA 240
      10 GTSSVASPPVQVFSTMLNRPRTLLASSPFHCLDKR-----EERLYLHLLLT 60
      241 CAEAVQENLSAADLVYKQIPLLAASOGAMRKVAAVYFGEALARRVRRPOPDSLLDA 300
      61 CANHVASGSLQANAAALQESHLASPDGDMQRIAAVFTETALNRILKSNP----GLYKA 116
      301 AFA-----DILAHN--FYESCPLYKFAHTANQALILEAFAGCRRVHVVDGKQGM 349
      117 LNAIQTRTNNVSEIHYRRLEFEMRPILKVSYLTLNRALIEAMEGEKVAHVYIDDASEPA 176
      350 QWPAALLQALALRPGGPPSFRITGYPPOPDETDALQGVGKLAQFAHTIRVDFQGLVA 409
      177 QWALALQAFNSRPGPPHRLRTGV-----HQKQVLEQMAHRLIEAEKLDLPQGFNPVVS 232
      410 -ATLADLEPMLQPGGEDPNEPEVIAVNSVFEHRLA----- 448
      233 RLDCINVEQLRV-----KTGEALAVSVSLQHLHTFLASDDDLMKRNCALRFQNNPSG 283
      449 -----QPGALEKVLGVRAVRPRI 467
      284 VDLQRLVIMSHGSAARENDMSNNNGYSPSGSASSLPPLPSSGRTDSFLNAINGLSFKV 343
      468 VTVVEOEANHNSTGLDRFTESLHYSTFDSLEGSSGGGSPSEVSSGMAAAPAAAGTDQ 527
      344 MVTVEQDSDNHGSTLMERLLESLTYTAALFCLD-----TVPRTSDRI 388
      528 VMSEVYLGROIQNVVACEGAEERTERHETLIGQWRNRNGNAGEVTVHLSNAYKQASTLLAL 587
      389 KVEKMLDGEELKNITISCEGFERRERHETLEKWSQRIDLAGGVNPLSYAMLAQARILLQ- 447
      588 FAGDGYKVEKEGCLTIGMHRPLIATSAW 619
      448 GCGFDGYRIKEEGCAVVICWQDRPLYSAMR 479

```







RESULT 13  
E86347  
hypothetical protein AAF87898.1 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: E86347  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maltli, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: E86347  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-593 <STO>  
A:Cross-references: GB:AE005172; NID:9454575; PIDN:AAF87898.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
Query Match 15.6%; Score 498; DB 2; Length 593;  
Best Local Similarity 29.6%; Pred. No. 1.8e-23;  
Matches 134; Conservative 76; Mismatches 169; Indels 74; Gaps 11;  
OY 184 STSSSSSSSSSGGARSVVEAAPPVAAANATPALPVVVVDQEGIRLVHALLAOAE 243  
DB 198 SPKSSSSADSNHSVSSKVVVSQATPK-----QIIISCAR 231  
OY 244 AVQENLSAAEALYKQIPDLAASOGAMRKVAAYFGEALARRV-----FRFR---POPD 294  
DB 222 ALSEKLEALSMVWELNQIVSIGDSPGRIAAVVEGLAARMAASGFIYALCKEPP 291  
OY 295 SSLLDAADADLLHAHFYSCPYLKPAHFTANAILLEAFAGCRHVAVDPGKIQGMQPAL 354  
DB 292 SDBRLAAQVVL-----FEVCPCKFKFGLAANGAILLEAKGEEVHIIDFDINQGNQMYTL 346  
OY 355 LQALALRGPGPSFRLNGVPPQDETD--ALQVGMKLAQAHITRVDFOYRGVLAAATL 412  
DB 347 IRSTIELTGKRRRLRLTGIDPESVQRSIGGRITGLRLEQLAEDNGVSFKFKAMPSTK- 405  
OY 413 ADLEPFML--QPEGEDNEPEVAVNSVFEHRL---LAQGLAEKVLGTVAAPR 466  
DB 406 SIVSSTLGCRP-----GETLIVFAFQQLHMPDESVTYVNRQBELHMKVSLNPK 456  
OY 467 IVTVVEQANHNSTGFLDRFTESLHYSTMDSLGSSGGSPSEVSSGAAAAPAAAGTD 526  
DB 457 LVTVEQDVNTTSPFPFRTIAYEYSAVESLD-----MTLPRESQER 501  
OY 527 QVMSVTVYGRQICNVVACGAEATERHEFTLQGNRNLGNAGETVHLGSMNAKQASTLLA 586  
DB 502 MNVERQCLARDIVNTVACGEEERIRYERAGKRAMMAGFNPMSAKVTNNITQNLIK 561  
OY 587 LFAGGDKYKVEKEGCLTGMHTRLPIATSAWR 619  
DB 562 Q-QYCNKTKKLEMEGELHFCWEKSLIVASAWR 593  
RESULT 14  
T51234  
scarcecrow-like protein 1 [imported] - Arabidopsis thaliana (fragment)  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 28-Jul-2000 #sequence\_revision 28-Jul-2000 #text\_change 28-Jul-2000  
C:Accession: T51234  
R:Pysh, L.D.; Mysicka-Diller, J.W.; Camilleri, C.; Bouchez, D.; Benfey, P.N.  
Plant J. 18, 111-119, 1999  
A:Title: The GRAS gene family in Arabidopsis: sequence characterization and basic expres

A:Reference number: Z25337; MUID:992722994  
A:Accession: T51234  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-352 <PYS>  
A:Cross-references: EMBL:AF036300; PIDN:AAD24403.1  
C:Genetics:  
A:Gene: SCL1  
Query Match 14.5%; Score 464; DB 2; Length 352;  
Best Local Similarity 31.4%; Pred. No. 1.2e-21;  
Matches 120; Conservative 67; Mismatches 147; Indels 48; Gaps 10;  
OY 235 ALVQIPLLAASOGAMRKVAAYFGEALARRV-----FRFR---POPDSSLLDAAPDL 305  
DB 2 SMVNLRLRIVSIGDSPGRIAAVVEGLAARMAASGFIYALCKEPPSDBRLAAQVL 61  
OY 306 LHAHFYSCPYLKPAHFTANAILLEAFAGCRHVAVDPGKIQGMQPALQALALAPGSP 365  
DB 62 -----FEVCPCKFKFGLAANGAILLEAKGEEVHIIDFDINQGNQMYTLIRIAELPKR 116  
OY 366 PSFRLTGVPPQDETD--ALQVGMKLAQFAHITRVDFOYRGVLAAATLADLEPML--Q 421  
DB 117 PRLRLTGIDDPESVQRSIGGLRITGLRLEQLAEDNGVSFKFKAMPSTK-SIVSPTLGCK 175  
OY 422 PEGEDPNEPEVAVNSVFEHRL---LAQGLAEKVLGTVAAPRIRIVTVVQEAH 477  
DB 176 P-----GETLLVNAFOLHMPDESVTYVNRQBELHMKVSLNPKLVVVBQDVT 226  
OY 478 NSGFLDRFTESLHYSTMDSLGSSGGSPSEVSSGAAAAPAAAGTDQVSEYVLGR 537  
DB 227 NTSPPFPRTIAYEYSAVESLD-----MTLPRESQERMANVERCLARD 271  
OY 538 ICNVVACGAEATERHEFTLQGNRNLGNAGETVHLGSMNAKQASTLLALFAGGQYVE 597  
DB 272 IVNIYACGEEERIEREYEAAGKRAMMAGFNPMSAKVTNNITQNLKQ-QYCNKTKK 330  
OY 598 EKEGCLTGMHTRLPIATSAWR 619  
DB 331 EEMGELHFCWEKSLIVASAWR 352  
RESULT 15  
T02736  
probable SCARECROW gene regulator [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 02-Feb-2001  
C:Accession: T02736; A84692  
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;  
submitted to the EMBL Data Library, August 1998  
A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.  
A:Reference number: Z14710  
A:Accession: T02736  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1336 <ROU>  
A:Cross-references: EMBL:AC005315; NID:93461834; PID:93461846  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Umayan, L.; Tallon,  
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: A84692  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1336 <STO>  
A:Cross-references: GB:AE002093; NID:93461846; PIDN:AC33232.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g29060; T914.14  
A:Map position: 2



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:24:50 ; Search time 17.09 Seconds  
(without alignments)  
1336.582 Million cell updates/sec

Title: US-09-485-529-7  
Sequence: 1 MKREYQDAGSGGGGGGMS.....TLGHTPLATSAWRLGP 623

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	152	4.8	375	1	SOX3_MOUSE
2	138	4.3	443	1	SOX3_HUMAN
3	136.5	4.3	1461	1	IE18_PRIVF
4	136	4.3	762	1	P115_CHICK
5	130.5	4.1	393	1	HXAA_HUMAN
6	130.5	4.1	507	1	IRX3_MOUSE
7	127.5	4.0	1446	1	IE18_PRIVK
8	127	4.0	399	1	HXAA_MOUSE
9	124.5	3.9	681	1	YDHE_SCHPO
10	124.5	3.9	825	1	ICP0_HSV2H
11	123.5	3.9	1676	1	APSA_EMENT
12	123.5	3.9	3038	1	TRIO_HUMAN
13	123	3.9	606	1	CENB_CRIGR
14	122.5	3.8	901	1	ALB0_MOUSE
15	119.5	3.7	2493	1	CYAA_USTMA
16	118.5	3.7	3567	1	ERY2_SACER
17	118	3.7	483	1	PREG_NEUCR
18	118	3.7	975	1	CUT1_CANFA
19	118	3.7	1395	1	CUT1_MOUSE
20	117.5	3.7	474	1	SOX4_HUMAN
21	117.5	3.7	915	1	ALB0_RAT
22	117.5	3.7	1059	1	CAPO_DROME
23	116.5	3.6	1001	1	ORL1_DROME
24	116.5	3.6	1093	1	AF17_HUMAN
25	116.5	3.6	1487	1	ICP4_HSVK
26	116	3.6	870	1	BCAL_HUMAN
27	116	3.6	889	1	CIKE_RAT
28	116	3.6	1294	1	CUT2_HUMAN
29	115.5	3.6	728	1	VIV_ORYSA
30	115	3.6	891	1	MAZ3_SCHCO
31	115	3.6	2442	1	CBP_HUMAN
32	114.5	3.6	1487	1	ICP4_HSVB
33	114.5	3.6	3421	1	TBCU_HSVB

34	114	3.6	500	1	BRN1_HUMAN
35	114	3.6	721	1	YR82_MYCTU
36	114	3.6	746	1	7UP2_DROME
37	114	3.6	1229	1	NI21_HUMAN
38	114	3.6	1298	1	ICP4_HSV1
39	113.5	3.6	1322	1	YAG3_YEAST
40	113.5	3.6	2038	1	F5H_DROME
41	113	3.5	995	1	Y103_YEAST
42	113	3.5	1367	1	AMTH_YEAST
43	113	3.5	2090	1	N214_HUMAN
44	113	3.5	2441	1	CBP_MOUSE
45	113	3.5	5179	1	MUC2_HUMAN

## ALIGNMENTS

RESULT 1  
ID SOX3\_MOUSE STANDARD; PRT; 375 AA.  
AC P53784;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE TRANSCRIPTION FACTOR SOX-3.  
GN SOX3 OR SOX-3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129;  
RA MEDLINE=96189340; PubMed=8625802;  
RA Collignon J., Sockanathan S., Hacker A., Cohen-Tannoudji M.,  
RA Norris D., Rastan S., Stevanovic M., Goodfellow P.N.,  
RA Lovell-Badge R.;  
RT "A comparison of the properties of Sox-3 with Sry and two related  
genes, Sox-1 and Sox-2.";  
RT Development 122:509-520(1996).  
RL -1- SUBCELLULAR LOCATION: NUCLEAR.  
RC -1- TISSUE SPECIFICITY: MAINLY IN THE DEVELOPING CENTRAL NERVOUS  
SYSTEM. EXPRESSED IN DEVELOPING UROGENITAL RIDGE.  
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X94125; CAA63845.1; -  
DR HSSP: Q05065; 1HRZ.  
DR MGD: MGI:98365; Sox3.  
DR InterPro: IPR000910; HMG\_12\_box.  
DR Pfam: PF00505; HMG\_box; 1.  
DR SMART: SM00398; HMG; 1.  
KM DNA-binding; Nuclear protein; Transcription regulation.  
FT DOMAIN 54 63  
FT DNA\_BIND 69 137  
FT HMG\_BOX 164 177  
FT DOMAIN 219 223  
FT POLY-ALA 223 223  
FT POLY-PRO 223 223  
FT POLY-ALA 252 259  
FT POLY-ALA 259 276  
FT POLY-ALA 276 276  
FT POLY-ALA 282 293  
SQ SEQUENCE 375 AA; 37857 MW; 2A8477A10A517FEF CRC64;  
Query Match 4.8%; Score 152; DB 1; Length 375;  
Best Local Similarity 20.4%; Prod. No. 0.023;  
Matches 69; Conservative 36; Mismatches 133; Indels 100; Gaps 8;



CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X15120; CAA33214.1; .

DR PIR: S04713; EDBEIF.  
KW Early protein; Transcription regulation; Trans-acting factor;  
FT DNA-binding; Phosphorylation; Nuclear protein.

FT DOMAIN 390 405 POLY-SER.  
FT DOMAIN 958 966 POLY-SER.  
SQ SEQUENCE 1461 AA; 149833 MW; 7F31E7ABE403B208 CRC64;

Query Match 4.3%; Score 136.5; DB 1; Length 1461;  
Best Local Similarity 22.1%; Pred. No. 1;  
Matches 158; Conservative 52; Mismatches 255; Indels 251; Gaps 30;

QY 8 AGSGGGGGGGSSSEDDKMWVSAAGE--GE-----EYDELLAALGYFVRAS 51  
DB 179 SSGSSSSSSSSPSPSAAPRRMSPARGDVGEPPAARPTPAPPAQAAVAANAARRGPS 238  
QY 52 DNADVAQKLEQLAMAMGCGVGAAPDDSFATHLATDTVHYNPTDLSWVESMSELNA 111  
DB 239 PASPAAGPVPS-----APGCGAPSGCGDGRHH-----HQHR-----EPLDEPA 279  
QY 112 ----PPPL-----PPAPQANASTSTSYT-----GSGGFDPD-- 140  
DB 280 ARRLDPRPLGANSVSSNPNSSSTTTVAVEPVANGPEKEDDGLAGDGG--APLOR 336  
QY 141 -----PSVDSSTSYALRPIPSAGATAPADLSADSVDRDKRMRTG 182  
DB 337 QRRRRAGGALRGHGFSSSSSGSDSLSPARSPARAPAAAAA---RRASSS 392  
QY 183 SSTSSSSSSSSSLGGARSSVVEAPVAAAANATPALPVVVVDTEAGIRLVHALLAC 242  
DB 393 SSSSSSSSSSSSEGEDEGCVPRGAPLARAAPPAPAAAPRSPASSASATSSAAS 452  
QY 243 EAVQOE-----NLISAALVLCQIPLA-----ASOGGA 270  
DB 453 PAPAPAPAPRRKRSTNNHLSLMDGPPPTDGPLTLPGEFWGSDDPPADGRVRYGA 512  
QY 271 -----MRKYAAYFGEALARRVFRFRPO-PDSSLDAAFADLHAHFY-ESCPY 316  
DB 513 GDSREGIMDEDDYROAAARY-RAAGRPVVFIPEDMGDSKQHVALVRLITYSGAAGEMSW 571  
QY 317 LKFAHTANQALILEAFAGC-RRYHVVDVFSIKOGMOPALLQALALRPGGPPSFRLTGVP 375  
DB 572 LONPRMQADDORENF--CORRYHA-----PHGHSITGSSVP 608  
QY 376 POPDEIDAL--QQVGKKLQAFHTIRV-----DFQYGLVA--- 409  
DB 609 PLPHIGDAAADOPMLALPHAVSAVMSRRYDRTOKFTLIQSLRAYADMAVYGRADPR 668  
QY 410 ---ATLADLEPMLQPEGEDEPNEEPE-----VIAVNSYFEMHRLI----- 447  
DB 669 AGATVTEALCARBARAFANAAGRGVPRRELADACVLAQCVLE--RLPCLRLPARARAP 726  
QY 448 --AOGALKEKVGTVARVPRIT--VTVEQDANHNSGTLDRFTESLHYIS----- 494  
DB 727 AALGPALCELEVPAALALDAIPGAGPAERQQAADSAVALVAVLVAVLVSVGARAAREA 786  
QY 495 -----TMF-----DSLEGGSSGGSPSEVSSG-----AAAAAPAG 524  
DB 787 AMTYAALAFRANVANAGARLAEAARPGAPAPAPGLPPLMPDEPGVLVAVPAPAPAG 842

RESULT 4

P115.CHICK  
ID P115.CHICK STANDARD: PRT; 762 AA.  
AC 098917;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE 115 KDA MELANOSOMAL MATRIX PROTEIN PRECURSOR.  
GN MPM115.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WHITE LEGHORN; TISSUE-Retinal pigment epithelium;  
RX MEDLINE=92020667; PubMed=1924173;  
RA Mochii M., Agata K., Kobayashi H., Yamamoto T.S., Eguchi G.;  
RT "Complete sequence and expression of a cDNA encoding a chicken  
RT 115-kDa melanosomal matrix protein.";  
RL pigment Cell Res. 4:41-47(1991).  
RN [2]  
RP CHARACTERIZATION.  
RC STRAIN-WHITE LEGHORN; TISSUE-Retinal pigment epithelium;  
RX MEDLINE=88311098; PubMed=3409326;  
RA Mochii M., Agata K., Kobayashi H., Yamamoto T.S., Eguchi G.;  
RT "Expression of gene coding for a melanosomal matrix protein  
RT transcriptionally regulated in the transdifferentiation of chick  
RT embryo pigmented epithelial cells.";  
RL Cell Differ. 24:67-74(1988).  
CC -1- FUNCTION: MIGHT BE REQUIRED FOR POLYMERIZATION OF MELANIN ONTO THE  
CC CORE STRUCTURE OF MELANOSOMES WITH ENZYMIC FUNCTION OF TYROSINASE.  
CC -1- SUBCELLULAR LOCATION: ON THE FIBROUS MATRIX STRUCTURE OF THE  
CC PREMELANOSOME.  
CC -1- TISSUE SPECIFICITY: SPECIFIC TO PIGMENTED EPITHELIAL CELLS AND  
CC MELANOCYTES. NOT EXPRESSED IN LENS, NEURAL RETINA, BRAIN, HEART,  
CC GIZZARD OR LIVER.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING THE REDIFFERENTIATION OF  
CC PIGMENTED EPITHELIAL CELLS (PEC).  
CC -1- PTM: GLYCOSYLATED.  
CC -1- SIMILARITY: BELONGS TO THE PMEL-17/NNB FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 PKD DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: D88348; BAA13589.1; .  
DR InterPro: IPR000601; PKD\_domain.  
DR Pfam: PF00801; PKD.1.  
DR SMART: SM00089; PKD.1.  
DR PROSITE: PSS0093; PKD.1.  
KW Signal; Glycoprotein; Repeat.  
FT SIGNAL 1 19  
FT CHAIN 20 762  
FT DOMAIN 223 323  
FT DOMAIN 441 532  
FT REPEAT 441 464  
FT REPEAT 465 488  
FT REPEAT 489 508  
FT REPEAT 509 532  
FT CARBOHYD 111 111  
FT CARBOHYD 115 115  
FT CARBOHYD 346 346  
FT CARBOHYD 651 651  
FT CARBOHYD 659 659  
SQ SEQUENCE 762 AA; 77356 MW; 172C8DBDAFDC7C6 CRC64;









DR SMART: SM00389: HOX. 1.  
 DR PROSITE: PS00027: HOMEBOX\_1; 1.  
 DR PROSITE: PS0071: HOMEBOX\_2; 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation; Alternative splicing.  
 FT DOMAIN 26 38 GLY-RICH.  
 FT DOMAIN 77 90 POLY-GLY.  
 FT DOMAIN 113 144 GIN/PRO-RICH.  
 FT DOMAIN 215 222 POLY-GLY.  
 FT DOMAIN 271 274 POLY-GLY.  
 FT DNA\_BIND 325 384 HOMEBOX.  
 FT VARSPIC 1 305 MISSING (IN ISOFORM 2).  
 FT VARSPIC 306 DSL -> MOO (IN ISOFORM 2).  
 SQ SEQUENCE 399 AA, 41415 MW, 7529624FC6057042 CRC64;

Query Match 4.0%; Score 127; DB 1; Length 399;  
 Best Local Similarity 22.1%; Pred. No. 0.76;  
 Matches 88; Conservative 37; Mismatches 157; Indels 116; Gaps 18;

QY 9 GSGGGGGGMSSEDKMYSAAG--EGEVDLLAALGKVRASMDVYQKLEOLEMA 66  
 ||| ||| : : : : :  
 DB 29 GSGAGGGGGYAHGCVYLPASDLRYGQSCGLFPALG-----SKRNEAPSP 76  
 ||| ||| : : : : :  
 QY 67 MGVGVGAGAAPDPSFATHATDTVHNPTDLSWVESMLS----- 107  
 ||| ||| : : : : :  
 DB 77 GGGGGGGSG-----LGPGHGIAPAPLDLMDAPRSCMEPPDPFPPOPQOO 127  
 ||| ||| : : : : :  
 QY 108 -ELNAPPPPLPAPQLNASTS---TVTSGGY--FDLPVSYSSTVALRP1---PSP 158  
 ||| ||| : : : : :  
 DB 128 QQQPPPPPPQPPQPPQANSCSFQNIKESSVCLVDADKCPKSAADLAFPPGPP 167  
 ||| ||| : : : : :  
 QY 159 AGTATADLSDSVRPKMKRTGSSSTSSSSSSSLGGAR--SSVEAPPVAAAANA 216  
 ||| ||| : : : : :  
 DB 188 DGCALGA---SSGVPPGYFRL--SQAYGAKFGSGGGGTEQLASFPAPQPP--GRGFD 241  
 ||| ||| : : : : :  
 QY 217 TPALPVVVDTOGATRLVHALLACAEOQENLSAAELVKQIP--LLAAGSGAMRK 273  
 ||| ||| : : : : :  
 DB 242 PPAL--ASSSTEAG-----KERVLDSTPPPTLVCTGGGSGQD 278  
 ||| ||| : : : : :  
 QY 274 VAAVFEALARRVRRPPQDSSILDAAFADLLHAHFES-----CPYLK- 318  
 ||| ||| : : : : :  
 DB 279 EEAHASSAAEEL---SPASENSKASPEKDSIGSGKGAANAWLAKSKKKRCPTTKH 335  
 ||| ||| : : : : :  
 QY 319 -----FAHFTANQALILEAFACCRVHVYDEGIK 346  
 ||| ||| : : : : :  
 DB 336 QTLLEKEFLFNMYLTRERLEI---SRVHLNDROVK 370  
 ||| ||| : : : : :

RESULT 9  
 YDHE\_SCHPO STANDARD; PRT; 681 AA.  
 ID YDHE\_SCHPO  
 AC 092359;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOHETICAL 73.3 KDA PROTEIN C6G9.14 IN CHROMOSOME I.  
 GN SPAC6G9.14.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NCBI\_TaxID=4896;  
 RN NCBI\_TaxID=4896;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Murphy L., Harris D., Bartell B.G., Rajandream M.A., Connor R.E.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBI databases.  
 CC -1- SIMILARITY: BELONGS TO THE PUMILIO/MPY5 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: 281317; CAB03616.1; -  
 DR InterPro: IPR001313; PWM.  
 DR Pfam: PF00806; PUF; 8.  
 DR SMART: SM00025; Pumilio; 8.  
 KW Hypothetical protein.  
 SQ SEQUENCE 681 AA, 73292 MW, 755A25C6A9A2D727 CRC64;

Query Match 3.9%; Score 124.5; DB 1; Length 681;  
 Best Local Similarity 20.2%; Pred. No. 2.1;  
 Matches 95; Conservative 74; Mismatches 135; Indels 167; Gaps 23;

QY 81 SFATHLATDTVHNPTDLSWVESMSELNAPPPPLPAPQLNASTSVTVGSGYFDLP 140  
 ||| ||| : : : : :  
 DB 113 SFTTSSSTTFPPPATMGG-----LNLSSP-----SSLYLPGSASY-QRS 155  
 ||| ||| : : : : :  
 QY 141 PSYDSSSIY---ALRPD-----SPGATAPADLSAD-----VRDPK 176  
 ||| ||| : : : : :  
 DB 156 NSKNSSASILQMTTLDIDIPILRRPQLSSYTPGPSTSRSSSSSLGNGPLIANNPS 215  
 ||| ||| : : : : :  
 QY 177 RMR-----TGGSTSSSSSSSLGGARS--SYVEAPPVAAA-----NATPALPVY 224  
 ||| ||| : : : : :  
 DB 216 ASKNFAFTSSSTNNSTSSMANGLQISKRSAFPLSNATSEFGENTLPPSLAAT 275  
 ||| ||| : : : : :  
 QY 225 V--DTQAGIRLVHALLACAEOQ-----NLAAE-----ALVKQIPLLAASOG 268  
 ||| ||| : : : : :  
 DB 276 ASYGSTDSSSNANTLTIATPTIPSPANTVGNPSADTPGNNVELIDDDSVSSLS 335  
 ||| ||| : : : : :  
 QY 269 GAMRKVAAVYGEAL-----RVFRFRPPQDSSILDAAFADLLHAHFESCPYLKFA 320  
 ||| ||| : : : : :  
 DB 336 SVAASLSLQNSNLSIFCKDHGCRYLQRLLEKKNQSHIDAVFA-----ETHPYLAV- 386  
 ||| ||| : : : : :  
 QY 321 HFTANQALILEAFAG--CRRVHVDFGIKQGMQMPALLQALARPGRPPRLTGVGPP 378  
 ||| ||| : : : : :  
 DB 387 -----LMDAFNGYLCOKL---FEHASEAONSTFQITAPKI--VIFSNMNG----- 429  
 ||| ||| : : : : :  
 QY 379 DETDALQOVCKLQAQFAHTTRVDFOYRGILVAATLADLEPMLQPEGEDEDENEBEVIAVN 438  
 ||| ||| : : : : :  
 DB 430 --TRALOKI-----ID----- 438  
 ||| ||| : : : : :  
 QY 439 SVFEMRLAOPGALKEVLTGTVRAVRPIVYVEQ--EANNHSGTELDRETE 488  
 ||| ||| : : : : :  
 DB 439 -----LVSSP---DQISCIYNALRPVNVLTFTDLNGNHVIQKLNKFSQ 479  
 ||| ||| : : : : :

RESULT 10  
 ICPO\_HSV2H STANDARD; PRT; 825 AA.  
 ID ICPO\_HSV2H  
 AC P28284;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICPO (VM118 PROTEIN).  
 GN RL2.  
 OS Herpes simplex virus (type 2 / strain HG52).  
 OC viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 NCBI\_TaxID=10315;  
 RN NCBI\_TaxID=10315;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92113549; PubMed=1662697;  
 RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;  
 RT Comparative sequence analysis of the long repeat regions and  
 RT adjoining parts of the long unique regions in the genomes of herpes  
 RT simplex viruses types 1 and 2.\*  
 RL J. Gen. Virol. 72:3057-3075(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Dolan A.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES ICPO PROTEIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D10471; BAA23427.1; -  
DR EMBL; 286099; CAB06760.1; -  
DR PIR; J01501; EDPEXD.  
DR InterPro; IPR001841; Znf\_fing.  
DR Pfam; PF00097; zf-C3HC4; 1.  
DR SMART; SM00184; RING; 1.  
DR PROSITE; PS00518; ZINC\_FINGER\_C3HC4; 1.  
KW Transcription regulation; Trans-acting factor; Activator; Zinc-finger;  
KW DNA-binding.  
FT DOMAIN 120 123 POLY-GLY.  
FT ZN\_FING 126 152 RING-TYPE.  
FT DOMAIN 166 271 POLY-GLY.  
FT DOMAIN 292 295 POLY-SER.  
FT DOMAIN 342 345 POLY-ALA.  
FT DOMAIN 385 389 POLY-SER.  
FT DOMAIN 395 400 POLY-GLY.  
FT DOMAIN 425 428 POLY-ALA.  
FT DOMAIN 590 627 POLY-SER.  
SQ SEQUENCE 825 AA; 81986 MW; SCEBI585853A274 CRC64;  
  
Query Match 3.9%; Score 124.5; DB 1; Length 825;  
Best Local Similarity 25.9%; Pred. No. 2.7;  
Matches 67; Conservative 22; Mismatches 95; Indels 75; Gaps 10;  
  
OY 11 SGGGGGCGSSSDKM-----MVSAAAGEGEVDELLALGYKVR 49  
DB 394 SGPGGGGLPQSSGRRAARPAAPRVSRPRAAAPVVSASADAAGPAPPAVNAHRAP 453  
OY 50 ASDMADV-----AQRLEOLEM--AMGCGVGA-----GAAPDDSFATHLATD 89  
DB 454 RSRMTQAOCTDQAQSLGRAGATDARGSGGPGVPRGTNTGGAAPH-----AAE 506  
OY 90 TVHYNP-----TDLSSWESMSELNAPPPPLP-----APQLNASTSTVTGSGGY 136  
DB 507 GAAPRRRRRGSDGSPAASSSSASSAAPPSPLAPQGVGAKRAAPRAPDSDS---GDRGH 563  
OY 137 FDL-----PPVSDDSSSIYALRPIPPAGATAPADISADSVDPKRMRTGSGSTSS 188  
DB 564 GPLAASAGAAPPSASPSSQAANVAASSSSASSSSSSSSSSSSSSSSSSSSSSSS 620  
OY 189 SSSSSSLGGGARSSVEAA 207  
DB 621 SSSASSAGAGGSVASAS 639  
  
RESULT 11  
APSA\_EMENT STANDARD; PRT; 1676 AA.  
AC Q00083;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE ANUCLEATE PRIMARY STERIGMATA PROTEIN.  
GN APSA.  
OS *Emmericella nidulans* (*Aspergillus nidulans*).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
CC Eurotiiales; Trichocomaceae; *Emmericella*.  
OK NCBI\_TaxID=5072;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-FGSC 4;  
RX MEDLINE=95164553; PubMed=7860626;  
RA Fischer R., Timberlake W.E.;  
RT "Aspergillus nidulans apsa (anucleate primary sterigmata) encodes a  
RT coiled-coil protein required for nuclear positioning and completion  
RT of asexual development.";  
RN J. Cell Biol. 128:485-498(1995).  
[2]  
RP CHARACTERIZATION  
RX MEDLINE=96025472; PubMed=9379904;  
RA Stelmann R., Slevens N., Fischer R.;  
RT "Nuclear traffic in fungal hyphae: in vivo study of nuclear migration  
RT and positioning in *Aspergillus nidulans*.";  
CC Mol. Microbiol. 25:757-769(1997).  
CC -1- FUNCTION: REQUIRED FOR NUCLEAR POSITIONING AND COMPLETION OF  
CC ASEXXUAL DEVELOPMENT.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.  
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X82289; CAA57733.1; -  
DR InterPro; IPR001849; PH.  
DR Pfam; PF00169; PH; 1.  
DR SMART; SM00233; PH; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
KW Coiled coil; Membrane.  
FT DOMAIN 51 127 COILED COIL (POTENTIAL).  
FT DOMAIN 193 359 COILED COIL (POTENTIAL).  
FT DOMAIN 408 453 COILED COIL (POTENTIAL).  
FT DOMAIN 1393 1504 PH.  
SQ SEQUENCE 1676 AA; 183671 MW; 8BDAL17F4212AEAD8 CRC64;  
  
Query Match 3.9%; Score 123.5; DB 1; Length 1676;  
Best Local Similarity 20.6%; Pred. No. 7.4;  
Matches 139; Conservative 73; Mismatches 222; Indels 241; Gaps 30;  
  
OY 19 GSSEDKMWSAAGEGEVDELLALGYKVRASDMADVAQRLEOLEMAMGCGVGAAP 78  
DB 936 GTSTIDADVSAISSEQIEPEVEPYERK--QDVAIYAEAVPEGPLSF-----VEQGTNT 988  
OY 79 DD--SFAT-----HLATDVHYNPDTLS----- 100  
DB 969 DDEVSFPALISSEVETPEVAPVRETKDDVEPYLSLDEQGTSTDEYFVSSTSETEPV 1048  
OY 101 -----WVESMLSELNAP-----PPP-----LPPAPQLNASTST 129  
DB 1049 EPIREAKEEAAVDDVASESTHPIFLTPPAYTEPTAPKLOEAVIEPPAPQLALSTVSS 1108  
OY 130 V-----TSGGVPFDLPVSDDSSSIYALRPIPPAGATAPADASA-----DSVRD-- 174  
DB 1109 VETPPVQYTPPVLLIPTPALDENTP-----PSVMASTAKATASAPPLVYVDNTDKG 1161  
OY 175 -----PKRMRTGSGSTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 203  
DB 1162 TADGLVTQONGVTLPGLAISGNAAPRRARS--GSSNQADQGAQTILSSKQIDQLIDRASV 1220  
OY 204 VEAAPVAAAANATPALPVYVVDTOEAGIRLVHALLACAEAVQENLSAELVYQIPL 263  
DB 1221 RPLSPDSDKINEMNSP-----FATPKARSRVPVPAASNLKRRGS 1263  
OY 264 AASOGAMRKVAAYGEGALARVFRFPQPOSSLDAFADLLHAFESCPYLFAHFT 323  
DB 1264 AASQASSVO-----IH-----PPLPADH-- 1281

```

OY 324 ANQALAEAFACGRVHVVDGKIGQGMQWPAALLQALAPGPPSFLTGVPQPEDTA 383
DB 1282 -----KEALMAAEKSIDORPASAGLMGPPLAPASAVRASSOOR-----ERTPMS-A 1328
OY 384 LQOVGNKLAQFAHTTRVD--FOYRGLVAATLADLEP-FMLQEGEEDPNEEPIAVANSV 440
DB 1329 LQVGSAKTTTSRASYRDSHMSRSSVSFASLEERFMNQP-----NPPFAQGYSTGTD 1384
OY 441 FEMHRLNLPQGLAEVYLG-----TVRAVRPRIVYVEQEAHNSTGTFIDRTESLHYTS 494
DB 1385 PRMIAAIQD-----TWIGELFMKRYTRRAVSGEL-----SNTRRRYFVWHPYTRL--YW 1432
OY 495 TWEDSLGGSSGGGSEVSGAA--PAAAGTD-OWMSEFYLGROIQVAVACEGAE 548
DB 1433 SEHDQOSAKSGSGRKSVSIEAVRVADNPFPGLHCKSLSEVSGRKIRPT-----AT 1487
OY 549 RTERHETLQWNRNL 563
DB 1488 TSORHET---WFNAL 1499

RESULT 12
TRIO_HUMAN STANDARD: PRT; 3038 AA.
AC 075962; Q13458;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE TRIPLE FUNCTIONAL DOMAIN PROTEIN (PTPRF INTERACTING PROTEIN).
GN TRIO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE-Fibroblast;
RX MEDLINE=96224308; PubMed=8643598;
RA Debant A., Serra-Pages C., Seipel K., O'Brien S., Tang M., Park S.-H.,
RA Streuli M.;
RA "The multidomain protein Trio binds the LAR transmembrane tyrosine
RT phosphatase, contains a protein kinase domain, and has separate rac-
RT specific and rho-specific guanine nucleotide exchange factor
RT domains";
RL Proc. Natl. Acad. Sci. U.S.A. 93:5466-5471(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RL Streuli M.;
RA Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP STRUCTURE BY NMR OF 1227-1407, AND MUTAGENESIS.
RX MEDLINE=99005194; PubMed=9790533;
RA Liu X., Wang H., Eberstadt M., Schnuchel A., Olejniczak E.T.,
RA Meadows R.P., Schkeryantz J.M., Janowick D.A., Harlan J.E.,
RA Harris E.A.S., Staunton D.E., Fesik S.W.;
RT "NMR structure and mutagenesis of the N-terminal Dbl homology domain
RT of the nucleotide exchange factor Trio.";
RL Cell 95:269-277(1998).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=99276567; PubMed=10341202;
RA Seipel K., Medley Q.G., Kedersha N.L., Zhang X.A., O'Brien S.P.,
RA Serra-Pages C., Hemler M.E., Streuli M.;
RT "Trio amino-terminal guanine nucleotide exchange factor domain
RT expression promotes actin cytoskeleton reorganization, cell migration
RT and anchorage-independent cell growth.";
RL J. Cell Sci. 112:1825-1834(1999).
RN [5]
RP FUNCTION: PROMOTES THE EXCHANGE OF GDP BY GTP. TOGETHER WITH
CC LEUCOCYTE ANTIGEN-RELATED (LAR) PROTEIN, IT COULD PLAY A ROLE IN
CC COORDINATING CELL-MATRIX AND CYTOSKELETAL REARRANGEMENTS NECESSARY
CC FOR CELL MIGRATION AND CELL GROWTH.
CC -! SUBUNIT: INTERACT TO FORM A COMPLEX WITH LEUCOCYTE ANTIGEN RELATED
CC PROTEIN.

```

```

CC -! ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -! TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, SKELETAL MUSCLE,
CC BRAIN, PANCREAS, PLACENTA, LIVER, KIDNEY AND LUNG.
CC -! DOMAIN: THE N-TERMINAL DBL/GEF DOMAIN SPECIFICALLY CATALYZES
CC NUCLEOTIDE EXCHANGE FOR RAC1, LEADING TO THE ACTIVATION OF JUN
CC KINASE AND THE PRODUCTION OF MEMBRANE RUFFLES. THE SECOND DBL/GEF
CC DOMAIN IS AN EXCHANGE FACTOR FOR RHOA AND INDUCES THE FORMATION OF
CC STRESS FIBERS.
CC -! PTM: SERINE PHOSPHORYLATED.
CC -! SIMILARITY: BELONGS TO THE RHO/RAC GEF FAMILY.
CC -! SIMILARITY: CONTAINS 2 DBL-HOMOLOGY DOMAINS (DB).
CC -! SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -! SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -! SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -! SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
CC -! SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SER/THR
CC FAMILY OF PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF091395; AAC43042.1; -
CC EMBL; U42390; AAC34245.1; -
CC HSSP; P04002; IATF.
CC MIM: 601893; -
CC InterPro: IPR001251; CRL_TRIO.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR003006; Ig_MHC.
CC InterPro: IPR003598; Ig_C2.
CC InterPro: IPR001849; PH.
CC InterPro: IPR000219; RhogEF.
CC InterPro: IPR001452; SH3.
CC InterPro: IPR002290; Ser_thr_kin_actsite.
CC InterPro: IPR002017; Spectrin.
CC pfam: PF00047; Ig_1.
CC pfam: PF00169; PH; 2.
CC pfam: PF00069; pkinase; 1.
CC pfam: PF00621; RhogEF; 2.
CC pfam: PF00018; SH3; 1.
CC pfam: PF00435; spectrin; 7.
CC SMART; SM00408; IGC2; 1.
CC SMART; SM00233; PH; 2.
CC SMART; SM00325; RhogEF; 2.
CC SMART; SM00516; SEC14; 1.
CC SMART; SM00326; SH3; 2.
CC SMART; SM00150; SPEC; 6.
CC SMART; SM00220; S_TKc; 1.
CC PROSITE; PS50003; PH_DOMAIN; 2.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS50002; SH3; 1.
CC Guanine-nucleotide releasing factor: Phosphorylation; Repeat;
CC Transferase; Serine/threonine-protein kinase; ATP-binding;
CC Immunoglobulin domain; SH3 domain; Alternative splicing.
CC REPEAT 252 359
CC REPEAT 479 585
CC REPEAT 819 925
CC REPEAT 1050 1157
CC DOMAIN 1240 1393
CC DOMAIN 1421 1532
CC DOMAIN 1597 1653
CC DOMAIN 1917 2091
CC DOMAIN 2098 2212
CC DOMAIN 2630 2707
CC DOMAIN 2737 2993
CC DOMAIN 656 659
CC POLY-GLN.

```

```
FT DOMAIN 1786 1791 POLY-SER.
FT DOMAIN 2233 2253 POLY-GLY.
FT DOMAIN 2486 2492 POLY-SER.
FT DISULFID 2637 2700 POTENTIAL.
FT NP_BIND 2743 2751 ATP (BY SIMILARITY).
FT BINDING 2766 2766 ATP (BY SIMILARITY).
FT ACT_SITE 2856 2856 BY SIMILARITY.
FT VARSPPLIC 2242 2308 GSGGGGGGCGAPSGSGHSGGSCGAPSTRSRPSRI
POPVRHHPVIVSSAASSQADKMS -> AAGVGAATAA
GPPVAAATVAAAPAAAAPPARAGAGPSPSLDTPPCW
SPLOPRAROKTRCO (IN ISOPORM 2).
FT VARSPPLIC 2309 3038 MISSING (IN ISOPORM 2).
FT MUTAGEN 1240 1240 E->A: 50% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1244 1244 T->A: 40% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1330 1330 N->A: NO CHANGE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1367 1367 V->A: 90% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1368 1368 Q->A: 80% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1369 1369 R->A: 80% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1371 1371 T->A: 80% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1372 1372 K->A: LOSS OF NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1375 1375 L->A: 40% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1378 1378 K->A: NO CHANGE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1379 1379 E->A: 30% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
SQ SEQUENCE 3038 AA: 341611 MW: 28620F3B513EB74B CRC64;
```

Query Match 3.9%, Score 123.5; DB 1; Length 3038;  
Best Local Similarity 23.7%; Pred. No. 16;  
Matches 58; Conservative 24; Mismatches 74; Indels 89; Gaps 10;

```
QY 4 EYODAGSGGGGGCGGMSSEDKMVAAGEGEVDLLAALGYKVRADMDVNAKQLQL 63
DB 2226 EYORNHSGGGGGGSGS-----GGGG----- 2247
QY 64 EMAMGCGVGA-----GAAPDSDFA--THLATDTVHNPTDLSWVESM-- 105
DB 2248 -----GSGGGGAPSGSGSHSGGSCGAPSTRSRPSRIPOPVRHHPVIVSSAASQAE 2303
QY 106 ---LSELNAPPEPLP---APOLNASTSTVTGSGGYFDLPSPVSSSSTIALRPIPSBA 159
DB 2304 ADKMSGTGPSPSLPPGAPAPAGPSAPSR-----RPGDADEGSEEREAEPIPKMK 2354
QY 160 GATAPADLSADSVRPKMKRMKRGSGSTSSSSSSSLG-----CGAASSVVEAPAPVAAAN 215
DB 2355 VLESRKKAANA-----SGSSPDAPAKDARASLTGLPKPRAGA--ASPLNSPLSS 2404
QY 216 ATPAL 220
DB 2405 AVPSL 2409
RESULT 13
CENB_CRIGR STANDARD: PRT: 606 AA.
AC P46988;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MAJOR CENTROMERE AUTOANTIGEN B (CENTROMERE PROTEIN B) (CENP-B).
GN CENPB.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_Taxid=10029;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary.
RX MEDLINE=96254058; Pubmed=8652663;
RA Bejarano L.A., Valdivia M.M.;
RT "Molecular cloning of an intronless gene for the hamster centromere
antigen CENP-B."
RL Blochim. Biophys. Acta 1307:21-25(1996).
CC -I- FUNCTION: INTERACTS WITH CENTROMERIC HETEROCHROMATIN IN
CC CHROMOSOMES AND BINDS TO A SPECIFIC SUBSET OF ALPHOID SATELLITE
CC DNA, CALLED THE CENP-B BOX. MAY ORGANIZE ARRAYS OF CENTROMERE
CC SATELLITE DNA INTO A HIGHER ORDER STRUCTURE WHICH THEN DIRECTS
CC CENTROMERE FORMATION AND KINETOCHORE ASSEMBLY IN MAMMALIAN
CC CHROMOSOMES (BY SIMILARITY).
CC -I- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: NUCLEAR.
CC -I- SIMILARITY: STRONG, WITH OTHER MAMMALIAN CENP-B.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U20951; AAB06494.1; -.
CC Chromosomal protein; Nuclear protein; DNA-binding; Centromere.
FT DNA_BIND 1 125
FT DOMAIN 404 471 GLU-RICH (ACIDIC).
FT VARSPPLIC 510 545 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 606 AA: 66407 MW: 63D0EEBE551D48E CRC64;
```

Query Match 3.9%, Score 123; DB 1; Length 606;  
Best Local Similarity 19.9%; Pred. No. 2.2;  
Matches 119; Conservative 60; Mismatches 226; Indels 194; Gaps 22;

```
QY 35 EEVDLLAALGYKVRASOM--ADVAKIQLLEEMAMGCGVGAAPDSDFAHTLATDTVH 92
DB 75 DKLEGLLAMPQOIRAAALPKVGIILKEKALRIAEELG-----MDPFTAS----- 119
QY 93 YNPDLSSWVESM-----LSELNAPPEPLPAPOLNASTSTVTGSGGYF-- 137
DB 120 -----NGMLDFRRRHGVACSGVTRSRARTSTPRAPAPAPAVPS--EGSGGSTPG 171
QY 138 -----DLPSPVDSSTIALRPIPSAGATAPADLSADSVRPKMKRMKRGSGSTSSSSSS 192
DB 172 WTRREOPPVSVEG--YASQVFS--ATETSLWYFLSDQASGLMGGDGTAKQATQR 224
QY 193 SLSGGGARSVVEAPVPAANAANATPALPVVVDTQEGIRLVHALLACAEVQOENLSA 252
DB 225 LSVLLCANBDGSEKPLPVAGKSAPR-----RASGGGLCDITYANKGCVT 270
QY 253 AEALVQIPLLAASOGAMRKVAAAYGEALARVFRFRPDPSSLLDAFADLHAHYE 312
DB 271 TQALAKYLKALDTRMAASRRVLLAGRLAOSL-----DTS-----GLRHVQL-- 314
QY 313 SCPYLKFAHTANQAILLEAFACGRVHYVDGFIKGM-----QWPALLQALALRGGPSF 368
DB 315 -----AFPPPGTVHPLERGVYQOVGVHQAQMLTKAMALEGQDPSG 356
QY 369 RUTGVGPQPDDETALQY--GWKLAQFAHTIRVDFOYRGVLAATLADLEPMLQPEGE 426
DB 357 LQUGL-----VEALHRYAAAMQAVEPA-DIATCFREAGFGCGANATITTT--SFSSEGE 407
QY 427 DNEPEEVIANSVEEMHRLLAQPALEKVLGTVAVPRIVTVEQEAHNHSGTFDLRF 486
DB 408 EEEEEE-----EEEEEE-----EEEEEEEGGEEBE 429
```

QY 487 TESLHYSTMPDLSGGSGGPSEVSGAAAPAAAGTDQVMSEVYLGRQICNVACEG 546  
 DB 430 EE-----EEGGEEGEGEGEE-----GEEVEE-----EG 455  
 QY 547 AERTREHETLGGWRNRLGNAGFEYVHLSNAYKQASTLIALFAGDGYVEKEGCLT 605  
 DB 456 DESDEEEEEESESSSEGLE-----AEDWAGVVEASGFGGYVGEACQPTL 507  
 RESULT 14  
 A180\_MOUSE STANDARD; PRT; 901 AA.  
 ID A180\_MOUSE 061548; 061547;  
 AC 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CLATHRIN COAT ASSEMBLY PROTEIN APL180 (CLATHRIN COAT ASSOCIATED PROTEIN).  
 DE APL180 (PHOSPHOPROTEIN F1-20) (91 KDA SYNAPTOSOMAL-ASSOCIATED PROTEIN).  
 DE SNAP91.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP MEDLINE=92300439; PubMed=1607933;  
 RA Zhou S., Sousa R., Tannery N.H., Lafer E.M.;  
 RT "Characterization of a novel synapse-specification protein. II. cDNA cloning and sequence analysis of the F1-20 protein.";  
 RL J. Neurosci. 12:2144-2155(1992).  
 CC -1- FUNCTION: ADAPTINS ARE COMPONENTS OF THE ADAPTOR COMPLEXES WHICH LINK CLATHRIN TO RECEPTORS IN COATED VESICLES. CLATHRIN-ASSOCIATED PROTEIN COMPLEXES ARE BELIEVED TO INTERACT WITH THE CYTOPLASMIC TAILS OF MEMBRANE PROTEINS, LEADING TO THEIR SELECTION AND CONCENTRATION. BINDING OF APL180 TO CLATHRIN TRISKELIA INDUCES THEIR ASSEMBLY INTO 60-70 NM COATS.  
 CC -1- SUBCELLULAR LOCATION: COMPONENT OF THE COAT SURROUNDING THE CYTOPLASMIC FACE OF COATED VESICLES IN THE PLASMA MEMBRANE.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: BRAIN. ASSOCIATED WITH THE SYNAPSES.  
 CC -1- DEVELOPMENTAL STAGE: DEVELOPMENTALLY REGULATED IN A PATTERN COINCIDENT WITH ACTIVE SYNAPTOGENESIS AND SYNAPTIC MATURATION.  
 CC -1- DOMAIN: POSSESSES A THREE DOMAIN STRUCTURE: THE N-TERMINAL 300 RESIDUES HARBOUR A CLATHRIN BINDING SITE, AN ACIDIC MIDDLE DOMAIN 450 RESIDUES, INTERRUPTED BY AN ALA-RICH SEGMENT, AND THE C-TERMINAL DOMAIN (166 RESIDUES).  
 CC -1- PTM: PHOSPHORYLATED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; M83985; AAA37587.1; -;  
 DR EMBL; M83985; AAA37586.1; -;  
 DR HSSP; P04002; IMA.  
 DR MGD; MGI:109132; Snap91.  
 DR InterPro; IPR001026; ENTH.  
 DR Pfam; PF01417; ENTH; 1.  
 DR SMART; SM00273; ENTH; 1.  
 KW Coated pits; Alternative splicing; phosphorylation.  
 FT DOMAIN 410 413 POLY-THR.  
 FT DOMAIN 535 539 POLY-ALA.  
 FT DOMAIN 547 550 POLY-ALA.  
 FT DOMAIN 659 664 POLY-SER.  
 FT DOMAIN 704 710 POLY-SER.  
 FT VARSPIC 715 719 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 901 AA; 91851 MW; 24A98FBACE8DB8B1 CRC64;

Query Match 3.8%; Score 122.5; DB 1; Length 901;  
 Best Local Similarity 22.7%; Pred. No. 3.9; Mismatches 212; Indels 181; Gaps 28;  
 Matches 132; Conservative 57;

QY 50 ASDMDVAOKLEOLEMAMOMGGVAGAA-----PDSFATH-----LATDHYVNPDLSSM 101  
 DB 337 AAPVSAKPSDDLIDQPPFSGAAGAAAPVPPSGATRAMGLDGLDGL-----AALSSV 392  
 QY 102 -VESMLSELNAPPPPLP-----PAPQNASTSTVTGSGGYFDLPSPVSSSYVALRPI 155  
 DB 393 PCFAPISDPFAPPEPSPPTTTTTPA-SASASTTAATVA-----VTTEVDFGDAFAAPG 445  
 QY 156 PSRA---GATAPAD-----LSADSVRPKRRKRTGSGSTSSSSSSSSSLGGARSSVE 205  
 DB 446 EAPAASEGATAPPTAPVAAALDACSGNDP-----FASEGSAGEAPELDFAMPPEP 499  
 QY 206 AAPVAAANAPALPVVVVDQEAGIRLVHA-LLACAAVOGENTSAEALVKQIP--- 261  
 DB 500 SAPVYPTASTAPPVATAPSPAPTVAAATTTTAAATTTATTSAAATTAAPRAL 559  
 QY 262 -----LLAASGGAMKVAAYFGELARRVFRPQPDSSLDAFADLLHAH 309  
 DB 560 DIFGDLFDSAPAEVAAAPKPDAPSIDLFGTDAFSSPPRGASVPESSLT---ADLLSYD 615  
 QY 310 FYESCPYLFHAFTANQALILEAFAGCRVHYVD-FGIRGOMQFALLQALALRPGGPSE 368  
 DB 616 AFPA-----PSPASTASPAEAESSG-----VIDLFGDARG----- 645  
 QY 369 RLTGVPQPDDELDALQYGMKLAQFAHTIRVDQYRGLVAATLADL-----EPFLQPE 423  
 DB 646 --SGASETP---APQAVSSSAS-----ADLLAGFGGSFMA--- 677  
 QY 424 GEDPPEPEVIAVNSVFEMHRLAOPGALKEVLTGTVRAVRPIVTVGQEAHNHNGTEP 483  
 DB 678 ----PSTTVTPAQN-----LQRP-SPEAFAFGTPSP----- 705  
 QY 484 DRFTESLHYSTWFDL-----EGSSGGGPSEVS---SGAAAPAAAGTDQVMSEVYL 534  
 DB 706 ---SSSSSPDSVFDLIGDLMPMTAPSGQPAVSPVSPAPMAASKGLSD-----L-755  
 QY 535 GRQICNVACEGAEERTREHETLGGWRNRLGNAGFEYVHLSN 576  
 DB 756 DSSLASLVGNLIGSTTSKGLDQW-----NAGEKRLTGGAH 792  
 RESULT 15  
 CYAA.USTMA STANDARD; PRT; 2493 AA.  
 ID CYAA.USTMA 01-FEB-1996 (Rel. 33, Created)  
 AC P49606;  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ADENYLATE CYCLASE (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE) (ADENYLATE CYCLASE).  
 DE UAC1 OR REM1.  
 GN Ustilago maydis (Smut fungus).  
 OS Ustilago maydis (Smut fungus).  
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;  
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.  
 OX NCBI\_TaxID=5270;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=518.  
 RX MEDLINE=95087882; PubMed=7995519;  
 RA Gold S., Duncan G., Barrett K., Kronstad J.W.;  
 RT "cAMP regulates morphogenesis in the fungal pathogen Ustilago maydis.";  
 RL Genes Dev. 8:2805-2816(1994).  
 CC -1- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER, cAMP.  
 CC -1- CATALYTIC ACTIVITY: ATP -> 3',5'-CYCLIC AMP + PYROPHOSPHATE.

CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.  
CC -1- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).  
CC -1- SIMILARITY: CONTAINS A PP2C-LIKE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: L33918; AAA57469.1; -  
DR InterPro: IPR001054; Guanylt\_cyclase.  
DR InterPro: IPR001611; LRR.  
DR InterPro: IPR003592; LRR\_out.  
DR InterPro: IPR003591; LRR\_tyr.  
DR InterPro: IPR003589; PP2C\_catalytic.  
DR InterPro: IPR001932; PP2C\_domain.  
DR InterPro: IPR000159; RA.  
DR Pfam: PF00211; guanylate\_cyc; 1.  
DR Pfam: PF00560; LRR; 15.  
DR Pfam: PF00481; PP2C; 1.  
DR SMART: SM00044; CYC; 1.  
DR SMART: SM00370; LRR; 6.  
DR SMART: SM00369; LRR\_tyr; 2.  
DR SMART: SM00332; PP2C; 1.  
DR SMART: SM00314; RA; 1.  
DR PROSITE: PS50125; GUANYLATE\_CYCLASES\_2; 1.  
KW Lyase; Repeat; Leucine-rich repeat; GMP synthesis; Magnesium.  
FT DOMAIN 759 763 POLY-ASP.  
FT DOMAIN 888 897 POLY-ALA.  
FT REPEAT 1084 1106 LRR 1.  
FT REPEAT 1108 1131 LRR 2.  
FT REPEAT 1132 1155 LRR 3.  
FT REPEAT 1157 1178 LRR 4.  
FT REPEAT 1179 1201 LRR 5.  
FT REPEAT 1202 1225 LRR 6.  
FT REPEAT 1227 1248 LRR 7.  
FT REPEAT 1249 1271 LRR 8.  
FT REPEAT 1273 1295 LRR 9.  
FT REPEAT 1314 1337 LRR 10.  
FT REPEAT 1338 1360 LRR 11.  
FT REPEAT 1361 1384 LRR 12.  
FT REPEAT 1386 1407 LRR 13.  
FT REPEAT 1408 1430 LRR 14.  
FT REPEAT 1432 1455 LRR 15.  
FT REPEAT 1509 1529 LRR 16.  
FT REPEAT 1533 1556 LRR 17.  
FT REPEAT 1557 1580 LRR 18.  
FT REPEAT 1581 1604 LRR 19.  
FT REPEAT 1606 1628 LRR 20.  
FT REPEAT 1633 1656 LRR 21.  
FT DOMAIN 1722 2001 PP2C-LIKE.  
FT DOMAIN 2002 2493 CATALYTIC.  
SQ SEQUENCE 2493 AA; 271979 MW; 106A872C3C1C5BF8 CRC64;

Query Match 3.7%; Score 119.5; DB 1; Length 2493;  
Best Local Similarity 23.2%; Pred. No. 21;  
Matches 66; Conservative 41; Mismatches 106; Indels 71; Gaps 12;

OY 6 ODAGSGGGGGG-----MGSSSEDKM-----WYSAAGGEEVDLLAAL----- 44  
DB 111 QDRSSGSGSGSGSRSPANSHSDSLPSPSTHSISPSFDE-DELROIMSHIPANQATSS 169  
OY 45 -----GYKVRASDMADVAQKLEOLEMAMGCGVGAGAAPDDSFATHLATDVIHYNPDTLS 99  
DB 170 SDGDYGRKAVQSANHQDISPFLQSE-----NAAPFSS-----SHSNRTSVN 210  
OY 100 SWESMSELNAPPPPLPPAPOLNASTSTSTVYGSGYFDLPPSVDDSSSIYALRPISPA 159  
DB 211 PSAASTASPSSTSAATRTTRPGGTNASTQYNTLDTSPGSDID-RPGLSSRSRQYSLRPQTPPS 269

OY 160 GATAPADLSAD-----SVRDP-----KRMRTGSGSTSSSSSSSSSIAGGARSSVVEA 206  
DB 270 ASTSTSLNGSKDTHASAVKKTNRNPFGLKKSSAHSNASSNHPTRHDI--GSVSSLSR 327  
OY 207 APPVAAANATPALFVYVVDIOEACIRLVHALLCAEAVQOENL 250  
DB 328 YGP-NAANVNPMPRPAPWLDN-----HCTLANSNSPSSASTL 362

Search completed: January 22, 2002, 16:30:28  
Job time: 338 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:24:25 ; Search time 40.9 seconds

(without alignments)  
2228.060 Million cell updates/sec

Title: US-09-485-529-7

Perfect score: 3194  
Sequence: 1 MKREYDAGSGGGGSGGMS.....TLCMHPPLATNSAMRLAGP 623

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.todent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3194	100.0	623	10	Q9ST59
2	2683	84.0	630	10	Q9ST48
3	2669	83.6	625	10	Q9MB96
4	1612	50.5	587	10	Q23642
5	1608	50.3	587	10	Q9SLH3
6	1604	50.2	587	10	Q23725
7	1594	49.9	532	10	Q23724
8	1594	49.9	533	10	Q9LOR8
9	1590	49.8	532	10	Q23643
10	1530.5	47.9	547	10	Q9SRP9
11	1511.5	47.3	511	10	Q9C8Y3
12	1495.5	46.8	662	10	Q65367
13	1450.5	45.4	523	10	Q9LFS3
14	1405.5	31.5	493	10	Q9A597
15	670	21.0	668	10	Q9FUZ7
16	606	19.0	617	10	Q9LW09
17	604.5	18.9	653	10	Q9M384
18	600.5	18.8	653	10	Q96304
19	599	18.8	428	10	Q9ZSP2

20	587.5	18.4	490	10	Q9LUL7	Q9LUL7 arabidopsis
21	585.5	18.3	819	10	Q9AVK4	Q9AVK4 pisum sativ
22	572.5	17.9	482	10	Q9LPR8	Q9LPR8 arabidopsis
23	558	17.5	526	10	Q9LPT0	Q9LPT0 arabidopsis
24	554	17.3	413	10	Q9LTH5	Q9LTH5 arabidopsis
25	551	17.3	658	10	Q9CAN3	Q9CAN3 arabidopsis
26	535	16.8	445	10	Q9ZWC5	Q9ZWC5 arabidopsis
27	527.5	16.5	584	10	Q9FL03	Q9FL03 arabidopsis
28	526	16.5	447	10	Q9ARF6	Q9ARF6 capsella ru
29	523.5	16.4	405	10	Q9FH21	Q9FH21 arabidopsis
30	512	16.0	313	10	Q9ZTB1	Q9ZTB1 zea mays (m
31	507	15.9	306	10	Q9XE53	Q9XE53 arabidopsis
32	498	15.6	593	10	Q9SD03	Q9SD03 arabidopsis
33	464	14.5	352	10	Q9XE51	Q9XE51 arabidopsis
34	462	14.5	1502	10	Q9LNX6	Q9LNX6 arabidopsis
35	445.5	13.9	1336	10	Q9L074	Q9L074 arabidopsis
36	444.5	13.9	542	10	Q9SCR0	Q9SCR0 arabidopsis
37	443	13.9	808	10	Q9XE58	Q9XE58 arabidopsis
38	439	13.7	610	10	Q9LTI5	Q9LTI5 arabidopsis
39	433.5	13.6	531	10	Q9SZF7	Q9SZF7 arabidopsis
40	429.5	13.4	284	10	Q9XE57	Q9XE57 arabidopsis
41	415	13.0	375	10	Q23566	Q23566 arabidopsis
42	402.5	12.6	325	10	Q9XE52	Q9XE52 arabidopsis
43	399.5	12.5	718	10	Q80933	Q80933 arabidopsis
44	399	12.5	287	10	Q9M0M5	Q9M0M5 arabidopsis
45	396	12.4	483	10	Q9SUP5	Q9SUP5 arabidopsis

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	623 AA.
Q9ST59	Q9ST59			
AC	Q9ST59			
ID	Q9ST59			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DR	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	GIBBERELLIN RESPONSE MODULATOR.			
GN	RHT-DIA.			
OS	Triticum aestivum (wheat).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;			
OC	Triticaceae; Triticum.			
OX	NCBI_TaxID=4565;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-99347734; PubMed-10421366;			
RA	Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,			
RA	Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F.,			
RA	Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;			
RT	"Green revolution" genes encode mutant gibberellin response			
RT	modulators."			
RL	Nature 400:256-261(1999).			
DR	EMBL: AJ242531; CAB51555.1; F2BAC3496D0A84F CRC64;			
SQ	SEQUENCE 623 AA; 65337 MW; F2BAC3496D0A84F CRC64;			
Query Match	100.0%; Score 3194; DB 10; Length 623;			
Best local Similarity	100.0%; Pred. No. 8.3e-199;			
Matches 623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 MKREYDAGSGGGGSGGMSSEDKMMVSAAGGEEYDELLAALGKVRASDADVAOKL 60			
DB	1 MKREYDAGSGGGGSGGMSSEDKMMVSAAGGEEYDELLAALGKVRASDADVAOKL 60			
QY	61 EOLEMMMGVGAAGAPDSEFATHLATDTVHYNPDLSSWVSMSLSELNAPPPLPPAP 120			
DB	61 EOLEMMMGVGAAGAPDSEFATHLATDTVHYNPDLSSWVSMSLSELNAPPPLPPAP 120			
QY	121 QUNASTSYTSGSGYFDLPSPVSSSSSYIALRIPSPAGATFAPDLSDSVDPKRMRT 180			
DB	121 QUNASTSYTSGSGYFDLPSPVSSSSSYIALRIPSPAGATFAPDLSDSVDPKRMRT 180			

[illegible]

RESULT	2
095ST48	
ID	Q9ST48
AC	Q9ST48;
DT	01-MAY-2000 (Tremblrel, 13, Created)
DT	01-MAY-2000 (Tremblrel, 13, last sequence update)
DT	01-MAR-2001 (Tremblrel, 16, last annotation update)
DE	GIBBERELLIN RESPONSE MODULATOR (FRAGMENT).
GN	
OS	Zea mays (Maize).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC	Panicoideae; Andropogoneae; Zea.
OX	NCBI_TaxID=4577;
OX	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=99347734; PubMed=10421366;
RA	Peng U., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,
RA	Flintham J.E., Beales J., Fish L.U., Worland A.J., Pelica F.,
RA	Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;
RT	'Green revolution' genes encode mutant gibberellin response
RT	modulators.";
RL	Nature 400:256-261(1999).
DR	EMBL: AJ242530; CAB51557.1; -.
FT	NON_TER
FT	630 630
Q	SEQUENCE 630 AA: 66028 MW: 3D56851726C51042 CRC64;

Query Match	84.0%;	Score 2683;	DB 10;	Length 630;
Best Local Similarity	85.8%;	Pred. No. 9.9e166;		
Matches 546; Conservative	27;	Mismatches 37;	Indels 26;	Gaps 13;

Qy	1	MKREYDAGGGGGGGGGSEDEKMYNSAA - AGCGEE -	-VDELLAALAGYVRASDADVA	57	
Db	1	MKREYDAGGS -	GGDMSSKKDKMMAAAAGAGEEDYDELLAALAGYVRSSDADVA	57	
Qy	58	QKLEQLEAMAMGCV - GAGAAPDDSFATHLATDVRHNP	TDLSWESMLSEINAPPL	116	
Db	58	QKLEQLEAMAMGCVGAGATADGPF	YSHLATDVRHNP	PDLSWESMLSEINAPPL	117

Qy	117	P--PAPDNNSTSYT---GSGGFEDLPBPVDSSTSYALRPIPSGAPADPLS	169
Dd	118	PPAPPAPL-ASTSYTSCGAAGAGTFDLPRAVDSSSTYALKPIPSYPAAS-ADPSI	175
Qy	170	DSVADPKMRGTGSSSTSSSSSSSSSGG-ARSSVBAAP--VAAANATPALPVVY	225
Dd	176	DSAREPKMRGTGSGSTSSSSSSSMGCGTRSSVBAAPVATQASADANG-PAPVVVY	234
Qy	226	DTQENGIRLVALLACAEVAOQENLSAAELVKOITPLAASOGAMKVAAYGELAR	285
Dd	235	DTQENGIRLVALLACAEVAOQENLSAAELVKOITPLAASOGAMKVAAYGELAR	294
Qy	286	VFRFRPODSSLDAAPADLLHAHFYESCPLYKFAHFTANOAILLEAFAGCRHVYVFGI	345
Dd	295	VYRFRPPDSSLDAAPADLLHAHFYESCPLYKFAHFTANOAILLEAFAGCRHVYVFGI	354
Qy	346	KQGMQWALLALALRGGPRSFRLTGVGRPOPETALDOOVGKLAQAFHTLRVDQYR	405
Dd	355	KQGMQWALLALALRGGPRSFRLTGVGRPOPETALDOOVGKLAQAFHTLRVDQYR	414
Qy	406	GLVATTLADLEPFMLQEGEDRPEREBEVAVANSVFEMHRLAQPGALEVLTGVRAVP	465
Dd	415	GLVATTLADLEPFMLQEGEDRPEREBEVAVANSVFEMHRLAQPGALEVLTGVRAVP	473
Qy	466	RIVYVVEQEAHNHNGSTPLDRTEBSLHYSTMPDLSBEGSSGGGSEVSSGAAAPAAAGT	525
Dd	474	RIVYVVEQEAHNHNGSTPLDRTEBSLHYSTMPDLSBEGSSGGGSEVSSGAAAPAAAGT	529
Qy	526	DOVMSVYLGRQIOINVVAECBAERTERHETLTGQMRNLGAGAGETVHLSGNAYKQASTLL	585
Dd	530	DOVMSVYLGRQIOINVVAECBAERTERHETLTGQMRNLGAGAGETVHLSGNAYKQASTLL	589
Qy	586	ALFAGGQGYKVEEKBGCLTJLGMHRRPLIATSAWMLA	621
Dd	590	ALFAGGQGYKVEEKBGCLTJLGMHRRPLIATSAWMLA	625

RESULT	3		
Q9MB96			
ID	Q9MB96	PRELIMINARY;	PRT; 625 AA.
AC	Q9MB96;		
DT	01-OCT-2000 (TREMBlrel. 15, Created)		
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)		
DT	01-OCT-2000 (TREMBlrel. 15, Last annotation update)		
DE	OSGAI.		
GN	OSGAI.		
OS	Oryza sativa (Rice).		
OC	Eurycotla: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	Ehartoideae; Oryzeae; Oryza.		
OX	NCBI_TaxID=4530;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. NIPONBARE;		
RX	MEDLINE=20179680; PubMed=10713441;		
RA	Ogawa M., Kusano T., Katsumi M., Sano H.;		
RT	"Rice gibberellin-insensitive gene homolog, OSGAI, encodes a nuclear-		
RT	localized protein capable of gene activation at transcriptional		
RT	level.";		
RL	Gene 245.21-29(2000).		
SD	EMBL; AB030956; BAA90749.1; -		
SD	SEQUENCE 625 AA; 65406 MW; 034FF02719D42E97 CRC64;		

Query Match	83.68;	Score 2669;	DB 10;	Length 625;
Best Local Similarity	85.98;	Pred. No. 7.9e-165;		
Matches 542; Conservative	24;	Mismatches	47;	Indels 18; Gaps 11;

```
QY      1 MKREYDAGGSGGGGGG--MGSSEDKMMVSAAGGEEVDELLAALGYKVRASDMADVAQ 58
      |||||:||||| ||| |||:|||||:|||||:|||||
Db      1 MKREYDAGGSSGGGSSADMGSCKDKVM-ACAAGEEEDVDELLAALGYKVRSSDMADVAQ 59
```

QY	59	KLEOLEAMGAGGAGAGAPDDSFATHLADTVYNTDLSNWSWESLSELMAPPLP	118
Db	60	KLEOLEAMGAGGAGAPADDCVSHLADTVYNTDLSNWSWESLSELMAPPLP <td>119</td>	119
QY	119	APOL--NASTSYVTGSG--GYFDLPSPVSSSSIVALARPIPSPAGATAPADLSADV <td>174</td>	174
Db	120	APPAARHASTSYVTGGGGSGFEFLPAAADSSSTVALRPLSLPVATADPS--AADSARD <td>178</td>	178
QY	175	PKRRRTGGSSSTSSSSSSSSSSISGGGA--RSSVYEAPV---AAAATPALPVVYDTQEA <td>230</td>	230
Db	179	TKRRRTGGSGTSSSSSSSSSSSSISGGGASGSVVEAAPPTOGAAAAANA--PAVPVVVDTOEA <td>237</td>	237
QY	231	GIRLVHALLACAEVVOENISAAALTKOJPLLAASGGAMRKVAAVEGALARRVFR <td>290</td>	290
Db	238	GIRLVHALLACAEVVOENISAAALTKOJPLLAASGGAMRKVAAVEGALARRVFR <td>297</td>	297
QY	291	PODPSLLDAAFADLLAHAFYESCPYLKFAHFTANOALIEAFAGCRHVHVVDGIRKGMQ <td>350</td>	350
Db	298	P-ADSTLLDAFAFDLLAHAFYESCPYLKFAHFTANOALIEAFAGCHRVHVVDGIRKGMQ <td>356</td>	356
QY	351	WPALLQALALRPGPPSPFRRLTGVPPOPDELDALQOYGMKLAOPAHITIRDFQIRGLVAA <td>410</td>	410
Db	357	WPALLQALALRPGPPSPFRRLTGVPPOPDELDALQOYGMKLAOPAHITIRDFQIRGLVAA <td>416</td>	416
QY	411	TLADLEPMLQPEGEENPNEPEPIAVNSYFEMRLLAOGALEKTVGTAVAPRLIVT <td>470</td>	470
Db	417	TLADLEPMLQPEGEENPNEPEPIAVNSYFELMRLLAOGALEKTVGTAVAPRLIVT <td>476</td>	476
QY	471	VEOBANHNSGTFDLRFTESLHYVSTMEDSLEGGSSGGGSPSEVSSGAAAPAAAGTDQVMS <td>530</td>	530
Db	477	VEOBANHNSGTFDLRFTESLHYVSTMEDSLEGGSS--GQALESPAG--GGGSTDQVMS <td>532</td>	532
QY	531	EVLIGRQICNVVACEGAEARTRPHETLQWRNRRLGNAGFEYVHIGSNAYKQASTLLAFAG <td>590</td>	590
Db	533	EVLIGRQICNVVACEGAEARTRPHETLQWRNRRLGRAGFEYVHIGSNAYKQASTLLAFAG <td>592</td>	592
QY	591	GDGKVEKEECGLTGMHTRPLINTSAMRLA 621 <td></td>	
Db	593	GDGKVEKEECGLTGMHTRPLINTSAMRYA 623 <td></td>	
RESULT	4		
ID	023642	PRELIMINARY;	PRT; 587 AA.
AC	023642;		
DT	01-JAN-1998 (TREMBlrel. 05, Created)		
DT	01-JAN-1998 (TREMBlrel. 05, last sequence update)		
DT	01-NOV-1999 (TREMBlrel. 12, last annotation update)		
GN	RGAL PROTEIN.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsits.		
OX	NCBI_Taxid=3702;		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=9379310; Pubmed=9237632;		
RA	Truong H.N., Caboche M., Daniel-Vedele F.;		
RT	"Sequence and characterization of two Arabidopsis thaliana cDNAs		
RT	isolated by functional complementation of a yeast gln3 gdh1 mutant."		
RL	FEES Lett. 410:213-218(1997).		
DR	EMBL; Y11336; CAAT2177.1; -		
DR	EMBL; 24145; Atachn;3051.2M445.		
DQ	SEQUENCE 587 AA; 64023 MW; D0A7A3C7A1PB5IEF CRC64;		

Query Match	50.5%	Score 1612;	DB10;	Length 587;
Best Local Similarity	53.7%	Pred. No. 1.8e-96;		
Matches 340; Conservative	82;	Mismatches 149;	Indels	62; Gaps
Oy	1 MKREYQDAG-----SGGGGCGMGSSSEDKMNV--SAAAGECEVDELLALGKYRASDMA	54		
	:		: ::   :	:   :

D	b	1	MKRHOHQFGSLNHHGTSSSSSSSISKOKMMVMKKEDGGGNMDELLAVLGYVRSEMA	60
O	y	55	DVAOKLEOLENAMGCVGAGAPDDSFATHLATDVYHPTDLSSWBSMSELNAPP	114
D	b	61	EVALKLEOLETMM-----SNVOEDGLSHLATDVYHNPSEIYSWLDMLSELN--PP	110
O	y	115	PLPAPQOLNATSSVTYGGSGYFLPSPVQ-----SSSYLARPIPSAGCTAPA-DLSAD	170
D	b	111	PLP-----ASSNOLDPVLPSPCEGPPADYDLKATPGNATYQFPADSSSS	157
O	y	171	SVRPFKRMRTGSGSSTS--SSSSSSSSLGCGARSSVVEAAPPVAAANAPALPVVVDTQ	228
D	b	158	SNQNKRLKSCSSPDSMTSTSTGTQIGVIGTGVTVTTTTTAAASTRS--VILVDSQ	215
O	y	229	EAGIRLVHALLACAEVQOEWLSAAELVYKOIPLLAOSGGAMRKVAATYGEMLARVER	288
D	b	216	ENGVLALVLAACAEVIOONMLTAEVLVYOIGCLAVSOAGMRKVAATYAEELARIRX	275
O	y	289	FRPODSSLDAAPADILLHAIFYESCPLYLKFPAHTANQALILEAFACRRVRHVVDGKOG	348
D	b	276	L--SPRQNDIDHCLSDTLQMHFYETCYLKFPAHTANQALILEAFEGKKRRHVHIDFSMNG	333
O	y	349	QMPWPLLOALALRPGGPPSEFLTVGPPQDETDALQOVGWLKAQFAHTITVDFOYRGIV	408
D	b	334	LQWPLMALALRGGGPPTEFLTIGPPAPDPSDHLHEVCSKTLAQLAEALHVEFEYRGFV	393
O	y	409	AATLADLEPFMLQDEGEDPNEBEVULVANSYFERMHLLAQALAEVLGVLTGVAANPRIV	468
D	b	394	ANSLADIDASML---ELRPS-DTEAAVANSYFELHLLLRPGGIEKVLVAVQKIRKIV	448
O	y	469	TJVEOEAHNNGTEFLDRPTESLHAYTSMFQPSLEGGSSGGGPPSEVSSGAAAPAAAGTDQV	528
D	b	449	TJVEDESHNHPVFLDRPTESLHAYTSLFPDSLEG-----VPNSQDKV	490
O	y	529	MSEYVLGQICNVVACBGAERTERHETLLGOMRNLGNAGETVHLSNAYKQASTLLALF	588
D	b	491	MSEYVLGQICNVLACBGPDEVERHETLLSQMGNFSSGSLAPAHGLSNAKQKQASMLSYF	550
O	y	589	AGGQGYKVEEKGCLTGLGMHRLPLIATSAMLA	621
D	b	551	NSGGYRVBESNGCLMGWHRPLITTSAMKLS	583

RESULT	5	
09SLH3		
ID	09SLH3	PRELIMINARY;
AC	09SLH3;	PRT; 587 AA.
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last annotation update)	
DE	PURINATE RGAL, GIBBERELLIN RESPONSE MODULATION PROTEIN.	
GN	ATG601570.	
OS	Arabidopsis thaliana (Mouse-ear cress).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;	
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsis.	
OX	NCBI_TaxID=3702;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CV, COLOMBIA.	
RX	MEDLINE=20083487; PubMed=10617197;	
RA	Lin X., Kaul S., Rounsley S.D., Shae T.P., Benito M.-I., Town C.D.,	
RA	Fuji C.Y., Mason T.M., Bowman J.L., Barnstead M.E., Feldhym T.V.,	
RA	Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., KO H., Moffat K.S.	
RA	Cronin L.A., Shen M., Vannken S.E., Unmayang L., Tallon L.J., Gill J.E.	
RA	Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,	
RA	Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A."	
RA	Salzberg S.L., Fraser C.M., Venter J.C.;	
RT	"Sequence and analysis of chromosome II of Arabidopsis thaliana."	
RL	Nature 402:761-768(1999).	
EMBL	AA005360; AAC67333.1;	
SO	SEQUENCE 587 AA; 64035 MW; FC92EF9408072AA CRC64;	

RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome II of *Arabidopsis thaliana*,";  
RL Nature 402:761-768(1999).  
DR EMBL; AC005560; AAC67333.1; -  
SQ SEQUENCE 587 AA; 64035 MW; FC92E7F9408072AA CRC64;

Query Match 50.3%, Score 1608; DB 10; Length 587;  
Best Local Similarity 53.6%, Pred. No. 3,3e-96;  
Matches 339; Conservative 82; Mismatches 150; Indels 62; Gaps 13;

```
QY 1 MKREYQDAG---SGGGGGGSGSSEDKMY--SAAAGEGEEVDELLAALGYKVRASDMA 54
D 1 MKRDHOFQGRLSNNGTSSSSSSSSSKDKMMYKKKEEDGCGNMDELLAVLGYKVRASDMA 60
QY 55 DVAOKLEOLEAMGCGVAGAGAAPDDSFATLADTATVYNNPDLSSWESMSELNAPPP 114
D 61 EYALKLEOLETMM-----SNVQEDGLSHLADTATVYNNPDLSSWESMSELN--PP 110
QY 115 PLPPAPQLNASTSVTSGGYFDLPSPVD--SSSIYALRPISPAGATAPA-DLSAD 170
D 111 PLP-----ASNGLDPLPSPICGFPASDYDLKVIQGNMAYGEPALDSSSS 157
QY 171 SVRDPKRRRTGGSSTS--SSSSSSSLGCGARSSVVEAAPPVAAAANATPALPVVVDQ 228
D 158 SNNQKRLKSCSSPDSSMTSTGTGIGVIGTGTGTTTTTTTAAESTRS--VILVDSQ 215
QY 229 EAGTRVLVALLACAAVOENLSAAELVKOIPLLASOGAMRKVAAYFGFALARVYR 288
D 216 ENGVALVHALMACAPALIOQNNLTLEALVKIGCLAVSOGAMRKVAAYFGFALARVYR 275
QY 289 FRPOPSSILDAAFADLLHAHFEESCPYLKFAHFTANOALIEAFAGCRRVHVVDGIRG 348
D 276 L--SPQNOIDHCLSDTLQMHFEYETCPYLKFAHFTANOALIEAFAGCRRVHVVDGIRG 333
QY 349 MOWPALLOALALRPGPSPFRITGVPQPDETDALOYGMKLAOPARTIRKDFYRGV 408
D 334 LQWPAIMQALALREGPPTFRITGIGPPAPDSDHLHEVGCLADLAIVHEFEYRGFV 393
QY 409 AATLADLEPFLQPGEDDPNEPEVIAVNSYFEMHRLLAOPGALEKVLGTVAARPRIV 468
D 394 ANSLADLDASML----ELRPS-DTEAVAVNSYFELHKLGRGGIEKVLGYVKQIKPIYF 448
QY 469 TVEQEANHNNGSTFLDRPTESLHYSTMFDSLEGSSGGGSPSEVSSGAAAAPAAAGTDQV 528
D 449 TVEQEENHNNGPVFLDRPTESLHYSTMFDSLEG-----VPNSQKIV 490
QY 529 MSEVYLQIOICNVACBGAERTERHETLQGNRNLGNAGFETVHLGSAVNAQASLTLLALF 588
D 491 MSEVYLQIOICNVACBGAERTERHETLSQWGNRNGSGLAPAHLSGNAFKQASMLLSVF 550
QY 589 AGDGYKVEEKEGCTLLGMHTRPLATISAMRLA 621
D 551 NSGCGYRVEESNGCLMGMHTRPLITTSAMKLS 583
```

RESULT 6  
023725 PRELIMINARY; PRT: 587 AA.  
AC 023725;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE GRS PROTEIN.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Peng J., Carol P., Richards D.E., King K.E., Cowling R.J.,  
RA Murphy G.P., Harberd N.P.;  
RL Genes Dev. 0:0-0(0).  
DR EMBL: Y15194; CAA75493.1;  
DR Mendel. 24071; Arabid.3051.24071.  
SQ SEQUENCE 587 AA; 64006 MW; F6F6C7738EE7DCA9 CRC64;

Query Match 50.2%, Score 1604; DB 10; Length 587;  
Best Local Similarity 53.4%, Pred. No. 6,1e-96;  
Matches 338; Conservative 82; Mismatches 151; Indels 62; Gaps 13;

```
QY 1 MKREYQDAG---SGGGGGGSGSSEDKMY--SAAAGEGEEVDELLAALGYKVRASDMA 54
D 1 MKRDHOFQGRLSNNGTSSSSSSSSSKDKMMYKKKEEDGCGNMDELLAVLGYKVRASDMA 60
QY 55 DVAOKLEOLEAMGCGVAGAGAAPDDSFATLADTATVYNNPDLSSWESMSELNAPPP 114
D 61 EYALKLEOLETMM-----SNVQEDGLSHLADTATVYNNPDLSSWESMSELN--PP 110
QY 115 PLPPAPQLNASTSVTSGGYFDLPSPVD--SSSIYALRPISPAGATAPA-DLSAD 170
D 111 PLP-----ASNGLDPLPSPICGFPASDYDLKVIQGNMAYGEPALDSSSS 157
QY 171 SVRDPKRRRTGGSSTS--SSSSSSSLGCGARSSVVEAAPPVAAAANATPALPVVVDQ 228
D 158 SNNQKRLKSCSSPDSSMTSTGTGIGVIGTGTGTTTTTTTAAESTRS--VILVDSQ 215
QY 229 EAGTRVLVALLACAAVOENLSAAELVKOIPLLASOGAMRKVAAYFGFALARVYR 288
D 216 ENGVALVHALMACAPALIOQNNLTLEALVKIGCLAVSOGAMRKVAAYFGFALARVYR 275
QY 289 FRPOPSSILDAAFADLLHAHFEESCPYLKFAHFTANOALIEAFAGCRRVHVVDGIRG 348
D 276 L--SPQNOIDHCLSDTLQMHFEYETCPYLKFAHFTANOALIEAFAGCRRVHVVDGIRG 333
QY 349 MOWPALLOALALRPGPSPFRITGVPQPDETDALOYGMKLAOPARTIRKDFYRGV 408
D 334 LQWPAIMQALALREGPPTFRITGIGPPAPDSDHLHEVGCLADLAIVHEFEYRGFV 393
QY 409 AATLADLEPFLQPGEDDPNEPEVIAVNSYFEMHRLLAOPGALEKVLGTVAARPRIV 468
D 394 ANSLADLDASML----ELRPS-DTEAVAVNSYFELHKLGRGGIEKVLGYVKQIKPIYF 448
QY 469 TVEQEANHNNGSTFLDRPTESLHYSTMFDSLEGSSGGGSPSEVSSGAAAAPAAAGTDQV 528
D 449 TVEQEENHNNGPVFLDRPTESLHYSTMFDSLEG-----VPNSQKIV 490
QY 529 MSEVYLQIOICNVACBGAERTERHETLQGNRNLGNAGFETVHLGSAVNAQASLTLLALF 588
D 491 MSEVYLQIOICNVACBGAERTERHETLSQWGNRNGSGLAPAHLSGNAFKQASMLLSVF 550
QY 589 AGDGYKVEEKEGCTLLGMHTRPLATISAMRLA 621
D 551 NSGCGYRVEESNGCLMGMHTRPLITTSAMKLS 583
```

RESULT 7  
023724 PRELIMINARY; PRT: 532 AA.  
AC 023724;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE GAI PROTEIN.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LANDSBERG ERECTA;  
RX MEDLINE=98051192; PubMed=9389651;  
RA Peng J., Carol P., Richards D.E., King K.E., Cowling R.J.,  
RA Murphy G.P., Harberd N.P.;  
RT "The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses";  
RL Genes Dev. 11:3194-3205(1997).



RESULT 9  
ID 023643 PRELIMINARY: PRT: 532 AA.  
AC 023643.  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE RGA2 PROTEIN.  
GN RGA2.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=97379310; PubMed=9237632;  
RA Truong H.N., Caboche M., Daniel-Vedele F.;  
RT "Sequence and characterization of two Arabidopsis thaliana cDNAs  
RT isolated by functional complementation of a yeast glr3 gdh1 mutant.";  
RL FEBS Lett 410:213-218(1997).  
DR EMBL; Y11337; CAA72178.1;  
DR Mendel; 24146; Arath;3051;24146.  
SQ SEQUENCE 532 AA; 58789 MW; 1AC719CA8072239 CRC64;

Query Match 49.8%; Score 1590; DB 10; Length 532;  
Best Local Similarity 54.2%; Pred. No. 4.3e-95;  
Matches 331; Conservative 72; Mismatches 106; Indels 102; Gaps 11;  
QY 22 EDKMWVSAAGEEVEDLLAALGYKVRASDADVAOKLEOLEMAMGCVGAGAAPDSS 81  
DB 11 DKKTMMNEEDDGNMDELLAVLGKVRSSSEADVAOKLEOLEMAMGCVGAGAAPDSS 63  
QY 82 FATHLATDTVHNPPTDLSWVESMLSELNAPPPLPPAPQLNASTSTVTGSGGYFDLPP 141  
DB 64 -LSQALATEVHNPAPALTYLWDSMLTDLNPP-----SNQOEDD- 93  
QY 142 SVDSSTSYALRPISPAATAPADLSADSVDPKRMRTGSSSTSSSSSLGCG--- 198  
DB 94 ---SSNAEDLKAIP-----GDALL-----NQFADISASSSSNOGGGDDT 129  
QY 199 -----ARSSVEAAPPPVAAAANATPALPVVVVDTQEAGIRLVHALLACAEVQOENL 250  
DB 130 YTTNKRKLKSNVVE-----TTTATAESTRHVVYLVDSQENGRLVHALLACAEVQOENL 184  
QY 251 SAAEALVKQIPLLAASQSGAMRKVAAYFGELARRVFRFPQPDSSLLDAFADLLHAHF 310  
DB 185 TVAEALVKGIGFLAVSQIGAMRQVATYFAEALARRIYRL--SPQSPIDHSLSDTLQMHF 242  
QY 311 YESCYLKAFAHTANQALILEAFAGCRVHVVDGIGKQGMQWPAALQALALRGGPSPRL 370  
DB 243 YETCYLKAFAHTANQALILEAFQGRKRVHVIDFSMSQIGQWPAALQALALRGGPSPRL 302  
QY 371 TGVGPPQPDDETDALQOVGMKLAQFAHTIRVDROYGLVAAATLADLEPRLQPEGDEPNE 430  
DB 303 TGIQPPAPDNDYDLEVEGCKLAHLAEALHVEFEYRGVANTLADIDASML---ELRPS- 357  
QY 431 EPEVAVNSVEFMRLAOPGALIEKVLCTVRAVRPRITVYVQEOANHNSGFLDRFTESTL 490  
DB 358 ELESAAVNSVEFLKHLKLRPAIDKVLGVNQIKPEIFTVVQESNNHSPFLDFTSTL 417  
QY 491 HYSYTMEDSLBGGSGGSPSEYSSGAAAAAPAAAGTDQVMSYVYLGKQICNVVACGAERT 550  
DB 418 HYSYTLFQSLSE-----VPSGQDKVMSYVYLLKQICNVVACGAPRV 459  
QY 551 ERHETLGMWRNRLNAGFETVHLGNSAYKQASTLLALFAGGQYVEKEGCLTGMWTR 610  
DB 460 ERHETLGMWRNRLNAGFETVHLGNSAYKQASTLLALFAGGQYVEKEGCLTGMWTR 519  
QY 611 PLIATSAWRLA 621

DB 520 PLIATSAWRLS 530  
RESULT 10  
ID 09SRP9 PRELIMINARY: PRT: 547 AA.  
AC 09SRP9.  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE RGA1-LIKE PROTEIN.  
GN T2IP5.13.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=CV. COLUMBIA;  
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,  
RA Rensing C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,  
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
RA "Arabidopsis thaliana chromosome III BAC T2IP5 genomic sequence.";  
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AC009895; AAF01590.1;  
SQ SEQUENCE 547 AA; 60493 MW; CAD18D951D95634 CRC64;

Query Match 47.9%; Score 1530.5; DB 10; Length 547;  
Best Local Similarity 51.3%; Pred. No. 3.2e-91;  
Matches 325; Conservative 83; Mismatches 107; Indels 119; Gaps 14;  
QY 17 GMGSSD---KMWVSAAGEEVEV-----DELLAALGYKVRASDADVA 57  
DB 4 GGETWDDPPKPLPASRGEGSPMDKKADDNNNSMDDDELLAVLGKVRSSSEADVA 63  
QY 58 QKLEOLEMAMGCVGAGAAPDSSFATHLATDTVHNPPTDLSWVESMLSELNAPPPLPP 117  
DB 64 QKLEOLEMAMGCVGAGAAPDSSFATHLATDTVHNPPTDLSWVESMLSELNAPPPLPP 108  
QY 118 PAPQLNASTSTVTGSGGYFDLPPSVDSSTSYALRPISPAATAPADLSA----- 169  
DB 109 -----ASDLDLTRS-----CYDRSE--YDLRAIPG-----LSAFPKEEYF 143  
QY 170 DSVRPPKRMRTGSSSTSSSSSSSLGCGARSSVVEAAPVAAAANATPALPVVVVDQ 229  
DB 144 DEASSKRIRLGSSWSSSDSESTRS-----VVLVDSQ 175  
QY 230 AGIRLVHALLACAEVQOENLSAAELVVKQIPLLAASQSGAMRKVAAYFGELARRVFR- 288  
DB 176 TGVRLVHALVACAEIHOENLNLADALVKRVCTLGSQAGAKYATYFAQALARRIYD 235  
QY 289 FRPQPD--SLDDAFAFADLLHAHFESCPYLKFAHTANQALILEAFAGCRVHVVDGIGIK 347  
DB 226 TVAEEDVCAAVNPSEFEVLEMHFESCPYLKFAHTANQALILEAVTTARRVVIDLGLNQ 295  
QY 348 GMQWPAALQALALRGGPSPSPRLTGVGPPQPDDETDALQOVGMKLAQFAHTIRVDQYK 407  
DB 296 GMQWPAALQALALRGGPSPSPRLTGVGPPQPDDETDALQOVGMKLAQFAHTIRVDQYK 355  
QY 408 VAAATLADLEPRLQPEGDEDPREEPEVLAVNSVFEHRLAOPGALIEKVLCTVRAVRPR 467  
DB 336 AAESLSDLEPEMF-----ETRPESETLVVNSVFEHRLAOPGALIEKVLCTVRAVRPR 409  
QY 468 VYVQEOANHNSGFLDRFTESTLHYSYTMEDSLBGGSGGSPSEYSSGAAAAAPAAAGTDQ 527  
DB 410 VYVQEOANHNSGFLDRFTESTLHYSYTMEDSLBGGSGGSPSEYSSGAAAAAPAAAGTDQ 453  
QY 528 VMSYVYLGKQICNVVACGAERTERHETLGMWRNRLNAGFETVHLGNSAYKQASTLLAL 587  
DB 454 VMSYVYLGKQICNVVACGAERTERHETLGMWRNRLNAGFETVHLGNSAYKQASTLLAL 513

```

QY 588 FAGGCGKYKEKGCCTTLCGHTRPILATSAMRLA 621
DB 514 YATGDBGYVEENDGCLMTGMOTRPLITTSAMKLA 547

RESULT 11
Q9C8Y3 PRELIMINARY: PRT: 511 AA.
ID 09C8Y3
AC 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE GIBBERELLIN REGULATORY PROTEIN, PUTATIVE.
T27F4.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Elgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malt R., Marshall A.,
RA Milttscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterbeck T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsids
thaliana."
RL Nature 408:816-820(2000).
DR EMBL: AC020665; AAG52171.1;
SQ SEQUENCE 511 AA; 56754 MW; 1B60071697C92A9F CRC64;

Query Match 47.3%; Score 1511.5; DB 10; Length 511;
Best Local Similarity 52.3%; Pred. No. 4,9e-90;
Matches 324; Conservative 65; Mismatches 118; Indels 113; Gaps 13;

QY 1 MKREYODAGSGGGGGGSGGSSSEDKMVAAGGEEVDELLAAGYKVRASMDADVAOKI 60
DB 1 MKREHHRRESSAGEG-SSSMTTVIKEBAAG---VDELLVYLGYKVRSSDMADVAHKI 54

QY 61 EOLEMAMGCVGAGAAPDDSFATHLATDTVHYNPTDLSWVESMSELNAPPPPLPPAP 120
DB 55 EOLEMYLGDG-----ISNLSDETVHYNPDSLGSWVESMSLDD--PTRIQEKP 100

QY 121 QLNASTSVTVTSGGYFDLPSPVSDSSSIYALRPIPSAGATAPADLSADSVDPKRMPT 180
DB 101 -----DSEYDLRAI--PGSAVYPRD--EHVYRKRRTKI 130

QY 181 GGSSTSSSSSSSSSLGGARSSVVEAAPVAAAANATPALPVVVVDVTQEGIRLVHALLA 240
DB 131 ESEILSTRS-----VVLDLSQEGVRLVHALLA 158

QY 241 CAEAYOENLSAAEALYKQIPILLASOGAMRKVAAYFGEALARRFRPRPODSSILDA 300
DB 159 CAAAYOONNKILADALVKVHGLASSOAGAMRRVATYFAEGLARIRYRIPRODVAL--S 216

QY 301 AFADLLAHFYESCPLYKFAHFTANQALIEAFAGCRRVHVVDGICOGQMOPALLQALAL 360
DB 301 AFADLLAHFYESCPLYKFAHFTANQALIEAFAGCRRVHVVDGICOGQMOPALLQALAL 360

```

```

DB 217 SPSDTLIQHFYESCPYLKFAHFTANQALIEVFATAEKVHVYIDLGNHGLQMPALLQALAL 276
QY 361 RRGCPSPFRLTGVGPPODETDALQOVKMLAQFAHTIRYDFQRLVAATLADLEPFI 420
DB 277 RPNGPDPDFRLTGIGYSLTD---IQEVGKMLGQLASTIGVNEFEKSIALNLSDLKPEMI 332

QY 421 QPGEEDPNEEPVIAVNSVFEMHRLLAQGALEKVLGYRAVRPRITVYVEQEAHNSG 480
DB 333 -----DIRGSEVAVNSVFELHRLLAHGSIDKFLSTKIRPIMTVVEQEAHNSG 386

QY 481 TFIIDRTESLHYSTMFDSLEGSSGSGPSEVSSGAAAPAAAGTDQVASEVYLGRQICN 540
DB 387 VFIDRTESLHYSTMFDSLEGSSGSGPSEVSSGAAAPAAAGTDQVASEVYLGRQICN 427

QY 541 VVACGAERTERHETVGMGNRLGNAGFETVHIGSNAYVQAOSTLLAFAGDGYKYEEK 600
DB 428 LVACEEDERERHETVGMGNRLGNAGFETVHIGSNAYVQAOSTLLAFAGDGYKYEEK 487

QY 601 GCCLTGMHTRPILATSAMRL 620
DB 488 GCCLTGMOTRPLITTSAMRI 507

RESULT 12
Q9C8Y3 PRELIMINARY: PRT: 662 AA.
ID 065367
AC 065367
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE RGAL PROTEIN.
GN RGA-LIKE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COL-0;
RA Sanchez-Fernandez R., Ardiles-Diaz W., van Montagu M., Inze D.,
RA May M.J.;
RT "Cloning of the Arabidopsis thaliana RGA-like gene, a putative member
of the WHIRL domain transcription factor family."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ224957; CA12242.1;
DR Mendel: 29006; Atath.3051;29006.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 1.
DR SMART: SM00320; WD40; 1.
DR PROSITE: PS50082; WD_REPEATS_2; 1.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
DR Repeat: WD repeat.
SQ SEQUENCE 662 AA; 73126 MW; 958AB73D42121CA8 CRC64;

Query Match 46.8%; Score 1495.5; DB 10; Length 662;
Best Local Similarity 51.7%; Pred. No. 7,4e-89;
Matches 320; Conservative 68; Mismatches 118; Indels 113; Gaps 13;

QY 2 KREYODAGSGGGGGGSGGSSSEDKMVAAGGEEVDELLAAGYKVRASMDADVAOKI 61
DB 153 QREHNRRESSAGEG-SSSMTTVIKEBAAG---VDELLVYLGYKVRSSDMADVAHKI 206

QY 62 OLEMAMGCVGAGAAPDDSFATHLATDTVHYNPTDLSWVESMSELNAPPPPLPPAP 121
DB 207 QLEMYLGDG-----ISNLSDETVHYNPDSLGSWVESMSLDD--PTRIQEKP 251

QY 122 LNASTSVTVTSGGYFDLPSPVSDSSSIYALRPIPSAGATAPADLSADSVDPKRMRTG 181
DB 252 -----DSEYDLRAI--PGSAVYPRD-----EHVYRKRRTKI 277

QY 182 GGSSTSSSSSSSSSLGGARSSVVEAAPVAAAANATPALPVVVVDVTQEGIRLVHALLAC 241

```

```

Db 278 RTEISELSSTRS-----VVLDISOETGVNLVHALLAC 310
OY 242 AEAVOQENLSAALVKQIPLLAASOGAMRKVAAYFGEALARVRFRPPODSSLLDAA 301
Db 311 AEAVOQNNLKADALVKVGLASSOAGAMRKVATYFAGLARIRYTRPRD--VASSS 368
OY 302 FADLHAHFESCPYLKFAHFTANOAILLEAFAGCRHVHVDFGIKQNMOPALLQALALR 361
Db 369 FSDLIQHIFYESCPYLKFAHFTANOAILLEVFATAKHAVIDLGHNHGLQWPAALQALALR 428
OY 362 PGPPSFRLTGVPQRPETDALQOVGKLAQFAHTIRVDFOYRGVLAATLADLEPPMLQ 421
Db 429 PNGPDRFLRTGTGYSLTP-----IOEVGKTLQGLASTIGVNEEFKSIANNLSDLPEML- 483
OY 422 PEGEEDPNEPEEPPVNVSVFEMHRLLAQGALEKVLGTAVRAPRTIVTVVEQANHNST 481
Db 484 -----DIRGLESVAVNVFELHRLAHPGSIDKFLSTIKSTIRPDIMTVBOEANNHNGTV 538
OY 482 FLDRTEESLHYSTWFDLSLEGSSGGSGPSEVSSGAAPAAAGTDQVNSEVYLGRQICNV 541
Db 539 FLDRTEESLHYSTSLFDSLE-----GPSQ-----DRVASELFLGRQIINTL 579
OY 542 VACBGAERTERHETLGQRRNLGNAGFETVHLGSAVYKQASTLLALFAGGSGYKVEEKG 601
Db 580 VACEGEDVHERHETLQRRNRFRGLGFKPVSIGSNAYKQASMLALYAGADGYNEENEG 639
OY 602 CLTLGWHTRPLIATSAWRL 620
Db 640 CLTLGWHTRPLIATSAWKRI 658

```

## RESULT 13

```

O9LF53 .PRELIMINARY: PRT: 523 AA.
AC O9LF53:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE RGA-LIKE PROTEIN.
GN K3M16_60.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
RA Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL391150; CAC01893.1;
SQ SEQUENCE 523 AA: 57326 MW: 0F6CE0BD13403C35 CRC64;

```

```

Query Match 45.4%; Score 1450.5; DB 10; Length 523;
Best Local Similarity 49.3%; Pred. No. 4.5e-86;
Matches 313; Conservative 76; Mismatches 115; Indels 131; Gaps 13;
OY 1 MKRREVDA-----GGSGGGGGGSGSSEDKMMVSAAGBEBVELLAALGY 46
Db 1 MKRSHQETSVBEAPASVKELENGCGGG-----DDNM-----DEFLAVLGY 42
OY 47 KVRASDMADVAKOLEBEMAGMGVGAAPDDSFATHLATDVHYHNPPTDLSVWESML 106
Db 43 KVRSSMDADVAKOLEBEMVLT-----SNDIASSSNAFEDYHYHNPPTDLSGNAQSM 93
OY 107 SELNAPPPLPPAPOLNASTSYTSGSGYFDLPPSVSSSYIALRPIPPACATAPAD 166

```

```

Db 94 SDLNYP-----DLDP-----NRICDLRPI-----TDDE 118
OY 167 LSADSVRDPKRMRTGSSSTSSSSSSSLGGARSSVVEAAPVVAANAANAPALPVVVD 226
Db 119 CCSNSNSNKRRLRGLPWCDSVTSSTRS-----VVL 149
OY 227 TOEAGIRLVHALLACAEVVOENLSAELVKQIPLLAASOGAMRKVAAYFGEALARRV 286
Db 150 IEETGVRLVQALVACAEVVOENLSLADALVKVGLLAASOGAMRKVAAYFAEALARRI 209
OY 287 FRFRPPODSSLLDAFADLHAHFESCPYLKFAHFTANOAILLEAFAGCRHVHVDFGIK 346
Db 210 YRI--HPSAALDPSEFETLQNNFYDSCPYLKFHFTANOAILLEVTSRRVHVVDLGIN 267
OY 347 QGMOPALLQALALALPGPPSFRLTGVPQRPDETDALQOVGKLAQFAHTIRVDFOYRG 406
Db 268 QGMOPALLQALALALRPGGPPSFRLTGVG--NFSNNEGIOELGKMLAQALQALGVFKKNG 325
OY 407 LVAATLADLEPPMLQPEGEEDPNEPEEPIAVNVSVFEMHRLLAQGALEKVLGTAVRAPR 466
Db 326 LTRERLSDLEPDMF-----ETRTSEETLVNVSVFELHRLAHPGSIDKFLSTIKLATVKAIRG 379
OY 467 IYTVVEQANHNHSGTFDRFTESLHYSTWFDLSLEGSSGGSGPSEVSSGAAPAAAGTD 526
Db 380 IYTVVEQANHNHGDVFLDRFENALHYSTSLFDSLE-----VYIPs---QD 423
OY 527 QVNSEVYLGRQICNVVACBGAERTERHETLGQRRNLGNAGFETVHLGSAVYKQASTLLA 586
Db 424 RYMSSEVYLGRQILNLYAVTBGSDRIERHETLQRRNRFRGLGFKPVSIGSNAYKQASML 483
OY 587 LFAGGSGYKVEEKGCLTLGWHTRPLIATSAWRLA 621
Db 484 LSGGSGDYRVEENDGSLMLAMQTKPLIAASAWKLA 518

```

## RESULT 14

```

O9AS97 .PRELIMINARY: PRT: 493 AA.
AC O9AS97:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE OSGAL.
GN P0707D10.30.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT clone:p0707D10.;
RT submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002910; BAB40172.1;
SQ SEQUENCE 493 AA: 51864 MW: 5FDB670EB8899492 CRC64;

```

```

Query Match 31.5%; Score 1005.5; DB 10; Length 493;
Best Local Similarity 43.1%; Pred. No. 2.6e-57;
Matches 222; Conservative 64; Mismatches 136; Indels 93; Gaps 11;
OY 113 PPPPLPPAPOLNASTSYTSGSGYFDLPPSVSSSYIALRPIPPAGATAPADLSASV 172
Db 24 PPPAAVADP-----DGVGYDPPAGADVDA-----ALPEFAAAPP----- 60
OY 173 RDPKRMRTGSSSTSSSSSSSLGGARSSVVEAAPVVAANAANAPALPVVVDTOEAGI 232
Db 61 -----CAPDAAAA-----VLAMREEEVAGI 82
OY 233 RLHVALLACAEAVO--QENLSAELVKQIPLLAASOGAMRKVAAYFGEALARRVFRFR 290

```



```

Db 83 RLVLHLMSCAGATFAGDHALASAOQLADSHALAAVSAAGISGRVAVHFTTALSRLFPSP 142
QY 291 PQPDSLLDAAFADLLHNFHFEPCYIKFAHFTANQATLLEAAGCRVAVVPGIKQGMQ 350
Db 143 VAPPT--DAEHAFYH--HEYACPYLKFHFTANQATLLEAAGCRVAVVPGIKQGMQ 199
QY 351 WPAILOALALRPGPPSFLTGVPQPPDETALOOVGMKLAOFHFTIVDPQRYGLVA 410
Db 200 WPAILOALALRPGPPSFLTGVPQPPDETALOOVGMKLAOFHFTIVDPQRYGLVA 259
QY 411 TLADLEPMLQPEGSEDPNEPEVIAVNSVFEMHRLAOP--GALEKVLGTVRAVPRP 467
Db 260 SLIDEPRMMLQ-----IAPGEAAVAFNSVLQHLRLGPDADQAPIDAVLDCVAIVPRKI 312
QY 468 VTVVQOEANHNSGTFLDRTESLHYSTMFDSLEGSSGGGSEVSSGAAAPAAATIQ 527
Db 313 FTVIEQEDHNTKGTGLDRTEALFYSAVFDSID-----AASASGAGN 356
QY 528 VASEVYLGROIQNVVACGEAERTERHETLGOVRNLGNAGFEVHLGSAVYQASTLLAL 587
Db 357 AMAEAYLQREIDYICGEGANRERHEPLSRWRDLTRAGLSAVPLGSAVNAHQARMLVGL 416
QY 588 FAGDGYKVEKEGCLITLGMHTRPLIATSAMRLAG 622
Db 417 PS-GEHGSVEADGCLTLGMHGRPLFSAASAMEAG 450

```

## RESULT 15

```

Q9FUZ7 09FUZ7 PRELIMINARY: PRT: 668 AA.
AC 09FUZ7

```

```

DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE SCARECROW.
GN SCR.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20407145; PubMed=10948251;
RA Lim J., Helariutta Y., Specht C.D., Jung J., Sims L., Bruce W.B.,
RA Diehn S., Benfey P.N.;
RT "Molecular analysis of the SCARECROW gene in maize reveals a common
RT basis for radial patterning in diverse meristems."
RL Plant Cell 12:1307-1318(2000).
DR EMBL; AF263457; AAG13663.1;
SQ SEQUENCE 668 AA: 71162 MW: 9287431046B2B621 CRC64;

```

Query Match 21.0%; Score 670; DB 10; Length 668;

Best Local Similarity 30.9%; Pred. No. 2e-35; Matches 202; Conservative 77; Mismatches 262; Indels 112; Gaps 19;

```

QY 8 AGSGGGGGGSGSGSDKMWVS-----AAAGEGEVD-----ELLAALGYKVA 50
Db 79 AAAAGVGGSGAASSASQALPQLHQLPFAFONHAREVDYRAHRAAHHQAGGEATA 138
QY 51 SDMDVYAOQKLEOLEAMGNGVGAAGAPDDSPATHLATDTVHYNPTDLSWVESML-SEL 109
Db 139 STTAAVWDGIIRI---IGSSG---GAAVSIQILINVRREIIRHPCNPGLASLELRLRSL 192
QY 110 NAPPPLPAPOLNASTSVTSGGYFLPRPSVSSSITVALRPPSPAGATAPADLSA 169
Db 193 AADPAPLPPPPQ-----POQHALLHGAPAAAPAGLTLPP-----PP 227
QY 170 DSVRDPKRMRTGSSSTSSSSSSSLGGGARSVVEAAPPVAAANATPALPVVV----- 224
Db 228 PPLPKRRHHPPCQQQQQEEPHAPQSPKAPTAETAAAAAQAQAAAAAAKAKERKEQ 287

```

```

QY 225 --VDQEGIRLVHALLACAEAVQENLSAEALYKQIPLLAASOGAMKVAAYGEAL 282
Db 288 RKRQREEGHLLTLTLQCAEAVNADNLDADQOTLIELATPFGTSTORVAATRAEAM 347
QY 283 ARRVFR-----FRPQDSSLLDAAFADLLHAF-----FYESCPLYKFAHFTANQATLE 330
Db 348 SARLVSSCLIVAPLPFGS-----PAARLHGVAAMAFQVFNCGISPPVKSHFTANQATLE 403
QY 331 ARAAGCRVAVVDPFGIKQGMQWPAILOALALRPGPPSFLTGVPQPPDETALOOVGMK 390
Db 404 AFEREERVHIIDIDIMQGLQWPGLFHILASRPGPPRVRLTGIGA-----SMEALFATGKR 459
QY 391 LAQFAHTIVDPQRYGLVAATLADLEPMLQPEGSEDPNE-----EEVIAVNSVFEMHRL 446
Db 460 LSDFADTLGLPFEFCAVA-----EKAGNVDPKLGTVRREAAVAVHMLH--HSL 505
QY 447 LAQPGALEKVLGTVRAVPRPIVVEQEANHNSGTFLDRTESLHYSTMFDSLEGSSG 506
Db 506 YDVTGSDSTIMLTIQRLAKKVITMVEQDLSH--SGSFLARFVEAIIHYTSALFDSLDSYGE 564
QY 507 GGPSEVSSGAAAPAAAGTDQVNSEVYLGROIQNVVACGEAERTERHETLGOVRNLGNA 566
Db 565 DSPER-----HVEEQQLSREIRNVLAAGGPART--GDYKFGSWREKLAQS 608
QY 567 GPETVHLGSAVYQASTLLALPAGDGYKVEKEGCLITLGMHTRPLIATSAMR 619
Db 609 GPPAASLAGSAAQAASLLIGMFP--SDGYTLVEENGALIKLGWKLCLITTSAMR 660

```

Search completed: January 22, 2002, 16:30:02  
Job time: 337 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:23:40 ; Search time 22.72 Seconds  
(without alignments)  
617.058 Million cell updates/sec

Title: US-09-485-529-7

Perfect score: 3194  
Sequence: 1 MKREYDAGSGGGGGGMS.....TLGHTPLATSAWRLGP 623

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:\*  
1: /cgn2.6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2.6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2.6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2.6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2.6/ptodata/2/1aa/PCYUS.COMB.pep:\*  
6: /cgn2.6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	1594	49.9	532	US-09-117-853-2	Sequence 2, Appli
2	589.5	18.5	282	US-09-117-853-8	Sequence 8, Appli
3	569.5	17.8	259	US-09-117-853-6	Sequence 6, Appli
4	406.5	12.7	221	US-09-117-853-4	Sequence 4, Appli
5	155.5	4.9	166	US-09-117-853-10	Sequence 10, Appli
6	127.5	4.0	1891	US-08-804-227C-12	Sequence 12, Appli
7	127.5	4.0	1891	US-08-804-198-6	Sequence 6, Appli
8	122	3.8	1580	US-08-804-227C-11	Sequence 11, Appli
9	122	3.8	1580	US-08-804-198-5	Sequence 5, Appli
10	121	3.8	4130	US-09-428-517-2	Sequence 2, Appli
11	120.5	3.8	1070	US-08-922-635-22	Sequence 22, Appli
12	119.5	3.7	3816	US-09-428-517-3	Sequence 3, Appli
13	119	3.7	1384	US-08-976-255-11	Sequence 11, Appli
14	119	3.7	4545	US-08-804-227C-14	Sequence 14, Appli
15	119	3.7	4550	US-08-804-227C-8	Sequence 8, Appli
16	119	3.7	4550	US-08-804-198-2	Sequence 2, Appli
17	118.5	3.7	3567	US-07-642-734C-4	Sequence 4, Appli
18	118.5	3.7	3567	US-08-439-009A-4	Sequence 4, Appli
19	117.5	3.7	1719	US-08-459-568-4	Sequence 4, Appli
20	117.5	3.7	1719	US-08-399-411-4	Sequence 4, Appli
21	117.5	3.7	1719	US-08-516-859A-4	Sequence 4, Appli
22	117	3.7	5588	US-09-036-987A-6	Sequence 6, Appli
23	117	3.7	5588	US-09-370-700-6	Sequence 6, Appli
24	116.5	3.6	1093	US-08-545-860D-55	Sequence 55, Appli
25	116.5	3.6	1093	PCR-US94-04496-55	Sequence 55, Appli
26	116	3.6	486	US-08-821-355A-8	Sequence 8, Appli
27	116	3.6	486	US-09-003-687A-8	Sequence 8, Appli

28	116	3.6	486	US-09-136-605-8	Sequence 8, Appli
29	116	3.6	511	US-08-821-355A-9	Sequence 9, Appli
30	116	3.6	511	US-09-003-687A-9	Sequence 9, Appli
31	116	3.6	511	US-09-136-605-9	Sequence 9, Appli
32	115.5	3.6	4928	US-09-036-987A-5	Sequence 5, Appli
33	115.5	3.6	4928	US-09-370-700-5	Sequence 5, Appli
34	114	3.6	1298	US-08-690-473-2	Sequence 2, Appli
35	114	3.6	1298	US-08-259-821A-2	Sequence 2, Appli
36	114	3.6	1298	US-08-843-659-2	Sequence 2, Appli
37	113.5	3.6	829	US-09-413-814-105	Sequence 105, App
38	113	3.5	2441	US-08-194-468-2	Sequence 2, Appli
39	113	3.5	2441	US-08-961-739-2	Sequence 2, Appli
40	112	3.5	1706	US-08-459-568-2	Sequence 2, Appli
41	112	3.5	1706	US-08-399-411-2	Sequence 2, Appli
42	112	3.5	1706	US-08-516-859A-2	Sequence 2, Appli
43	111.5	3.5	1147	US-08-131-365B-38	Sequence 38, Appli
44	111.5	3.5	1147	US-08-668-123-38	Sequence 38, Appli
45	111.5	3.5	3729	US-08-804-227C-4	Sequence 4, Appli

#### ALIGNMENTS

RESULT 1  
US-09-117-853-2  
; Sequence 2, Application US/09117853  
; Patent No. 6307126  
; GENERAL INFORMATION:  
; APPLICANT: Harberd, Nicholas P  
; APPLICANT: Peng, Jinyong  
; APPLICANT: Carol, Pierre  
; APPLICANT: Richards, Donald E  
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana  
; FILE REFERENCE: 620-45  
; CURRENT APPLICATION NUMBER: US/09/117,853  
; CURRENT FILING DATE: 1998-08-12  
; EARLIER APPLICATION NUMBER: PCT/GB97/00390  
; EARLIER FILING DATE: 1997-02-12  
; EARLIER APPLICATION NUMBER: GB 9602796.6  
; EARLIER FILING DATE: 1996-02-12  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 532  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-117-853-2

Query Match 49.9%; Score 1594; DB 4; Length 532;  
Best Local Similarity 54.3%; Pred. No. 9,7e-125;  
Matches 332; Conservative 71; Mismatches 106; Indels 102; Gaps 11;

22 EDKMTYSAAGSEEEDELLAAGYVRASDMADVAKOLEEMAMGSGVGAAPDD5 81  
11 DKRTKMNEDDSDNGDELLAAGYVRSEMDADVAKOLEEVM-----SNVEDD- 63  
82 FATHLTDTYVYAPPTLSSVSESMSELNAPPPLPAPDOLNASTSYVSGGYFDLPP 141  
64 -LSQLATEYHYVYAPDLYWDSMLDLNPP----- 93  
142 SYDSSSYIALRPISPAAGTAPADLSADSVDPKRMRTGGSSSTSSSSSSSLGGG--- 198  
94 ---SSNAEYDLKAIIP-----GDAITL-----NOFAIDSSSSSSNOGGGCP 129  
199 -----ARSSVYEAAPVAAANATPALPYVYVDTQEGFTLVHALLACAAVQOENL 250  
130 YTTNKRKCKSGNGVE-----TTTATSESTRHYVLDVDSQENGVALHALACAAVQKENL 184  
251 SAAEALVKOIPLLAASQAGMKRVAAVFGAALARRFFRPPQDSSILDAFAFDLHAHF 310  
185 TVAEALVYKQIGFLAVSQIGAMKRVATYFAALARRIYRL--SPSQSPIDHSLSDTLQMHF 242  
Db



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	12.7%	Score 406.5;	DB 4;	Length 221;
Best Local Similarity	40.3%	Pred. No. 2.7e-26;		
Matches 106; Conservative	30;	Mismatches 48;	Indels 79;	Gaps 8;

[illegible]

RESULT 5  
US-09-117-853-10

```

Sequence ID: Application US/09117853
Patent No. 6307126
GENERAL INFORMATION:
APPLICANT: Harberd, Nicholas P
APPLICANT: Peng, Jinrong
APPLICANT: Carol, Pierre
APPLICANT: Richards, Donald E
TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
FILE REFERENCE: 620-45
CURRENT APPLICATION NUMBER: US/09/117,853
CURRENT FILING DATE: 1998-08-12
EARLIER APPLICATION NUMBER: PCT/GB97/00390
EARLIER FILING DATE: 1997-02-12
EARLIER APPLICATION NUMBER: GB 9602796.6
EARLIER FILING DATE: 1996-02-12
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10
LENGTH: 166
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-117-853-10

```

Query Match	4.98;	Score 155.5;	DB 4;	Length 166;
Best Local Similarity	25.78;	Pred. No. 1.5e-05;		
Matches 61; Conservative	27;	Mismatches 50;	Indels 99;	Gaps 10

```

0Y  DVAKLEIDLEAKMGMGVGNAADDSFAHIIADPHYNYPTDLSNWSVMSLSTLNP 114
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  DVAKLEIDLEAKMGMGVGNAADDSFAHIIADPHYNYPTDLSNWSVMSLSTLNP 76
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
0Y  FLPPAPOLMASTSTVTGSGGYFDLPSPVSDSSSYALRLPIPSFAGATAAPADLSADV 174
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db  SSNAEYDLKAI-----GDAIL- 93
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |

```

QY	175	PKRRRTGGSTSTSSSSSSSLGCGARRSSVEAPVAAAANATPALVVDVDTQEAIRL	234
		:    :	
		-----	
Db	94	---NQPAIDSSSSNNGGGG-----DYTTNKR-	118
QY	235	VHALLACAEAVQENLSAAELVYQIPLLAASGGCAKMKVAAATYFGEA-LARRVYRFR	250
		:    :	
		-----	
Db	119	---LKCSNGVETTTTAAES-TTHVSMLT-----RRTYCVSTRRMLDKLRRR	165

RESULT 6  
US-08-804-227C-12

Sequence 12, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rosteck, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804.227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1891 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-227C-12

Query Match	4.0%	Score 127.5;	DB 2;	Length 1891;
Best Local Similarity	21.8%	Pred. No. 0.12;		
Matches 154;	Conservative 72;	Mismatches 290;	Indels 189;	Gaps 31;

QY	30	AAGGEEVDELTALAKKVRASD----	NADVAQKTEOLEEMAMGCV-----	GAGAPD	79
Db	347	ALGDTEIGALLATTYG--RERVGDPLMTGSLNSTNGHQAAGVCGVKKVQAMHGSIP-			404
QY	80	DSFATHLATDTVAHN-PTDLSNWE----	SMSELNAPPPLPPAPOLNASTSTVTGSGG		135
Db	405	-----RLTHVDAPSSKYMMSGAVELLTEGRSNPRRVERVRR--AVSAFVGSTNA			454
QY	136	YFDLPSPVSDSSSIYALRPITS-PAGATAPAD--LSADSVADPKRMKTGSGSTSSSSSS			192
Db	455	HVYLEEAFAVEGSHGDDGPGDPREDVATGTPPLWYLSARS-REALRGQAGRLAALARGRT			513
QY	193	SSLGGARSSVVEA-----	PPAAANANATPALPVP		223
Db	514	EGTGGGS-GLVVPADADICYSIATTRETLERHAVALVOENRAGEDIALAALAGRPESEVIT			572
QY	224	VVDIQEAGIRLVHALLACAEAVQOENTSAAEALVKQIPLLAASGGAMRKVAAYFGELA			283

```

Db 573 GVARGRGI-----AFICSGGQGR--LGAGRELGRPPVFA-----DALDEIAAEFDAHLE 622
QY 284 RR-----VERFRPODSSILD-----AAFA-----DLHAHFYESCYPY 316
Db 623 RPLLSVFAEPATPDALDRTDYTOPALFAVEFALFRLESQGLVPDVLVGH---SIGG 679
QY 317 LKFAH-----FTANQAILLEAFAGCRVHVVDGIGQMGWMPALLQALALRP----- 362
Db 680 LVAHVAGVFSAADARLVASGRILMRALPEGAMAAVQATEREAALEPVAAGAVYAA 739
QY 363 -GGPSPFRITGVGPPQDETDALQOVGWKLAQFAHTTIVDQYRGVLAATLADLEPFMLQ 421
Db 740 VNGPQALVLSG-----DEAAVLAAGELARGRTRK----- 771
QY 422 PEGEDPNEEPEEVIANSVFEMHRLAOPGALKEVLTGVRAVRPRIVTVVEQEAHNHSGT 481
Db 772 -----LRVSHAFISPRMDALADFAVADTVDHARLPVSEV-----TGD 813
QY 482 FLDR-----FTESLHYSTMFDSLEGSSG-----GPSEVSSGAAAPAAAGTD 526
Db 814 LADAQQLTDPGYWTRQVQPVRFADAVRTASARDAATFIELGPDVAVLCGMAEESLAAD 873
QY 527 QVMS-EVYLGROIQCNVACEGAERTERHETLQWRNRLGNAGFETVHLSNAYKO----- 580
Db 874 VVFAPALRGREGDTVLRAAASAVVRGAGL-DWAALYGGTGARRTDLPTVAFQHSRYWL 932
QY 581 --ASTLLALFAGGDGK---VEEKEGCLTLGWHTRPLIATSAMRL 620
Db 933 APASAAVAPATPAAPSVRSVPFAEDDGLMAAVHAGD-VASAAARL 976

RESULT 7
US-08-804-198-6
: Sequence 6, Application US/08804198
: Patent No. 5945320
: GENERAL INFORMATION:
: APPLICANT: Burgett, Stanley G.
: APPLICANT: Kustoss, Stuart A.
: APPLICANT: Rao, Nagaraja R.
: APPLICANT: Richardson, Mark A.
: APPLICANT: Rosteck, Paul R., Jr.
: TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PAUL R. CANTRELL 1138
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: Macintosh
: OPERATING SYSTEM: Macintosh 7.0
: SOFTWARE: Microsoft Word 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,198
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: CANTRELL, PAUL R.
: REGISTRATION NUMBER: 36,470
: REFERENCE/DOCKET NUMBER: P9113
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-3885
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1891 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: MOLECULE TYPE: peptide
: US-08-804-198-6

```

```

Query Match 4.0%; Score 127.5; DB 2; Length 1891;
Best Local Similarity 21.8%; Pred. No. 0.12;
Matches 154; Conservative 72; Mismatches 290; Indels 189; Gaps 31;

QY 30 AAGECEYDELLAALGKVRSD---NADVAKLEQLEMMGMGV-----GAGAPD 79
Db 347 ALGPIEGALLATYG--KERGDPLMLGSLKSNIGHQAAGAGGVLIKVOAMRHGSLP- 404
QY 80 DSFATHLATDVHYN-PTDLSWYE-----SMLEINAPPPLPPQOLNASTSTVTSGGC 135
Db 405 -----KTLHADBSKYMASGAVELLTEGRSMRREVRERRA-AVSAFVGSTNA 454
QY 136 FYDLPVSDSSSIYALRPDS-PAGATAPAD--LSADSVDPKRMRTGSSSTSSSSS 192
Db 455 HVYLEEAPVEAGSEHGDPGDPDPAVAGPLPVLVSARS-REALRGQAGRIALAARQRT 513
QY 193 SSLGGARSSVVEAA-----PVVAAAANATPALPVV 223
Db 514 EGTGGGS-GLVVPADIGYSLATTRELEHRAVALVDENRTAGEDLALALAGRTPESVYT 572
QY 224 VVDIOEAGIRLVHALLCAEAVQENLSAAELVKQIPLLAASOGAMRKVAAVFGAALA 283
Db 573 GVARGRGI-----AFICSGGQGR--LGAGRELGRPPVFA-----DALDEIAAEFDAHLE 622
QY 284 RR-----VERFRPODSSILD-----AAFA-----DLHAHFYESCYPY 316
Db 623 RPLLSVFAEPATPDALDRTDYTOPALFAVEFALFRLESQGLVPDVLVGH---SIGG 679
QY 317 LKFAH-----FTANQAILLEAFAGCRVHVVDGIGQMGWMPALLQALALRP----- 362
Db 680 LVAHVAGVFSAADARLVASGRILMRALPEGAMAAVQATEREAALEPVAAGAVYAA 739
QY 363 -GGPSPFRITGVGPPQDETDALQOVGWKLAQFAHTTIVDQYRGVLAATLADLEPFMLQ 421
Db 740 VNGPQALVLSG-----DEAAVLAAGELARGRTRK----- 771
QY 422 PEGEDPNEEPEEVIANSVFEMHRLAOPGALKEVLTGVRAVRPRIVTVVEQEAHNHSGT 481
Db 772 -----LRVSHAFISPRMDALADFAVADTVDHARLPVSEV-----TGD 813
QY 482 FLDR-----FTESLHYSTMFDSLEGSSG-----GPSEVSSGAAAPAAAGTD 526
Db 814 LADAQQLTDPGYWTRQVQPVRFADAVRTASARDAATFIELGPDVAVLCGMAEESLAAD 873
QY 527 QVMS-EVYLGROIQCNVACEGAERTERHETLQWRNRLGNAGFETVHLSNAYKO----- 580
Db 874 VVFAPALRGREGDTVLRAAASAVVRGAGL-DWAALYGGTGARRTDLPTVAFQHSRYWL 932
QY 581 --ASTLLALFAGGDGK---VEEKEGCLTLGWHTRPLIATSAMRL 620
Db 933 APASAAVAPATPAAPSVRSVPFAEDDGLMAAVHAGD-VASAAARL 976

RESULT 8
US-08-804-227C-11
: Sequence 11, Application US/08804227C
: Patent No. 5876991
: GENERAL INFORMATION:
: APPLICANT: Dehoff, Bradley S.
: APPLICANT: Kustoss, Stuart A.
: APPLICANT: Rosteck, Paul R., Jr.
: APPLICANT: Sutton, Kimberly L.
: TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: THOMAS G. PLANT 1501
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1580 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-804-227C-11

```

```

Query Match          3.8%; Score 122; DB 2; Length 1580;
Best Local Similarity 20.7%; Pred. No. 0.26; Indels 208; Gaps 33;
Matches 135; Conservative 78; Mismatches 230;

```

```

QY 6 ODAGSGGGGGGMSSEDEKMMVSA--GGEVEDELLAAL 44
DB 301 QD-GASNGILTAGSPAQOVIREALDAGLTPADVDVAEHGCTGPLDPIAGALMATY 359
QY 45 GYKVASD---MADVAQKLEQLEMAMGMGV-----GAGAAPPDSEFTHLATDTVHYN 94
DB 360 GHE-RTGDPMLGSLKSNIGHTQAAAGVATKMYLALRHGELP-----RTLH-- 406
QY 95 PDLSSWVE-----SMSELNAPPPPLPAPQLANASTSTVTGSGCYF---DLPPSVDS 145
DB 407 ASTASSRIEMDAGAVELLDEAR-PWPRRAEGPRRAGISSFGISGTNAHLVIEEPPA--- 462
QY 146 SSSIVALLRPIPS---PAGATAPADLSADSVDPKRMRTGSSSTSSSSSSSSLOGGAR 200
DB 463 -----REPPEAAQPPAPATTVPLPSAAGAR-----SLREQARRLAHLAGHEE 506
QY 201 SSVVEAAPVVAANAATPAPLVVVVDTOGAGIRLVHALLACAEAVQOENLSAAEA----- 255
DB 507 ITAADAAARSAATTTRALSHRASVLDADRRA---LIDRLTALAEDEKRDPGVTGVEAGSGRP 563
QY 256 -----LVKQIPILASOGGAMRKVAAYFEBALARVFRFRP----- 291
DB 564 PVFVPPGGSSQMTGMAEELLDRAPVFRKAEBECARALAAHLDMSVL-DVLRDAPGAPPID 622
QY 292 -----QPPSSLLDAFAADLLHAH-----FYESCPLYLKFH----- 321
DB 623 RADVVOPTLFTMMVSLAALMESHGVRPAVVGHSOGCEITAAHAAGLSLDDARVIAERS 682
QY 322 -----FTANQATLEAFACGRVHVVDFTKQGMQ-WPALLQALALRPGGPPSFRLTGVP 375
DB 683 RLMKRLAGNGMLSYVAPADRV-----RELMEPVAEMESVAAV---NGPASTVAVG--- 730
QY 376 PQPDELDALQOYG-----WKLA--QFA-HPIRVDFOYR-----GLVATLADL 415
DB 731 ----DARALEEGGRLSAAGVLRMPLAGVDFGHSFOYE-QFRAELLDTLGTVRPTAARL 785
QY 416 EEPFMT-----QPPGEEDPNEEPEVIAVSVFEMHRLAQGALAEKVLGTVRAVFRIV 468
DB 786 -PEFSVTAAHEPEG-----LDAAY-WYRNRRREVEFASTLRTL--LREGHR 829
QY 469 TYVEDEANINSGTFLDRFTESLHYSTMPDSLGGSSGGGSEVSSGAATA 519
DB 830 TEVEMGPHLLAAIDEVAEGVHATLATLHRGSGGIDRRRSSVGAFA 880

```

```

RESULT 9
US-08-804-198-5
; Sequence 5, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R. A.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostek, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1580 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-804-198-5

```

```

Query Match          3.8%; Score 122; DB 2; Length 1580;
Best Local Similarity 20.7%; Pred. No. 0.26; Indels 208; Gaps 33;
Matches 135; Conservative 78; Mismatches 230;

```

```

QY 6 ODAGSGGGGGGMSSEDEKMMVSA--GGEVEDELLAAL 44
DB 301 QD-GASNGILTAGSPAQOVIREALDAGLTPADVDVAEHGCTGPLDPIAGALMATY 359
QY 45 GYKVASD---MADVAQKLEQLEMAMGMGV-----GAGAAPPDSEFTHLATDTVHYN 94
DB 360 GHE-RTGDPMLGSLKSNIGHTQAAAGVATKMYLALRHGELP-----RTLH-- 406
QY 95 PDLSSWVE-----SMSELNAPPPPLPAPQLANASTSTVTGSGCYF---DLPPSVDS 145
DB 407 ASTASSRIEMDAGAVELLDEAR-PWPRRAEGPRRAGISSFGISGTNAHLVIEEPPA--- 462
QY 146 SSSIVALLRPIPS---PAGATAPADLSADSVDPKRMRTGSSSTSSSSSSSSLOGGAR 200
DB 463 -----REPPEAAQPPAPATTVPLPSAAGAR-----SLREQARRLAHLAGHEE 506
QY 201 SSVVEAAPVVAANAATPAPLVVVVDTOGAGIRLVHALLACAEAVQOENLSAAEA----- 255
DB 507 ITAADAAARSAATTTRALSHRASVLDADRRA---LIDRLTALAEDEKRDPGVTGVEAGSGRP 563
QY 256 -----LVKQIPILASOGGAMRKVAAYFEBALARVFRFRP----- 291
DB 564 PVFVPPGGSSQMTGMAEELLDRAPVFRKAEBECARALAAHLDMSVL-DVLRDAPGAPPID 622
QY 292 -----QPPSSLLDAFAADLLHAH-----FYESCPLYLKFH----- 321

```

Db 623 RADVQPTLFTMVNSLAALWESHGVRPAAYVGHSGEIAAHAAGALSDDAARYIARS 682  
QY 322 -----FTANQALILEAFACCRVHVDFGIKQOMQ-WPALDQALALRPGSPSFRITGVGP 375  
Db 603 RUMKRLAONGMLSMALADRY-----RELMEPMARMYAAY--NGPASYVAG--- 730  
QY 376 POPBETDALQOVG-----MKLA--QPA-HTIRVDFQYR-----GLVATTLADL 415  
Db 721 ----DARALEERGRSLAAGVLRMPLAGVDFAGHSPOYE-QFRADLLDTLGTVRPTAARL 785  
QY 416 EPPML-----QPEGEDPDEEPEVIVANSVFEMHRLLAQPGALEKVLGYRAVRPRIV 468  
Db 786 -PEFSTVTAHAHEPG-----LDAAV-WYRMKREPEVERASTLRVL--LRGHR 829  
QY 469 TVVEEAAHNSGTFLDRFTESLHYXTMDESLGSSSGGSPSEVSSGAAA 519  
Db 830 TFEVGGPRLGAAIDVAAEAGVHATATLTLHGSGGLDRFRSSVGAAFA 880

RESULT 10  
US-09-428-517-2  
; Sequence 2, Application US/09428517  
; Patent No. 6251636  
; GENERAL INFORMATION:  
; APPLICANT: Betlach, Mary C.  
; APPLICANT: Shah, Sanjay Krishnakant  
; APPLICANT: McDaniel, Robert  
; APPLICANT: Tang, Li  
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 30062-20029.00  
; CURRENT APPLICATION NUMBER: US/09/428.517  
; CURRENT FILING DATE: 1999-10-28  
; EARLIER APPLICATION NUMBER: 60/120.254  
; EARLIER FILING DATE: 1999-02-16  
; EARLIER APPLICATION NUMBER: 60/106.100  
; EARLIER FILING DATE: 1998-10-29  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4150  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant  
; OTHER INFORMATION: Oleandolide PKS  
US-09-428-517-2

Query Match 3.8%; Score 121; DB 4; Length 4150;  
Best Local Similarity 20.7%; Pred. No. 1.3;  
Matches 147; Conservative 69; Mismatches 209; Indels 286; Gaps 35;

QY 14 GCGGN---GSSEDKM-----MVSAAA-----GEGEVDLLAAL--GYKV 48  
Db 3199 GCGGNVSGVGLSAERRTMLDTYGGVSAVAANGPSSIVSGDAQALDELLAGCEREGRVA 3258  
QY 49 R-----ASMDADVAQKLEOLEMAMGKGVGAGAAPDPSFTHLATDVIHNPDLSSWV 102  
Db 3299 RRVVDYASHAQMOLDELLEAL-----ADVTPODSSVPEFTV-----ADWL 3304  
QY 103 ESMSELNAPPPLPPAPQLASTSTVTGSGGVFDLPPVSSSSSIYALRPIPSAGAT 162  
Db 3305 D-----TTALDAGYF----- 3315  
QY 163 APADLSADVDPKRMRTGSGSTSSSSSSSSSGGARSSVEEAPVYAAAANATPALPV 222  
Db 3316 -----TNLEETVAFQ-----EAVEGLVAGMGAEVCECPHVLVGLITETL-- 3356  
QY 223 VVVDTEAGITLVHALLACAFAVOOENLSAAEALVKQIPL--LAASOGAMRKV--AAVF 278  
Db 3357 ---DFPDAAVALSLSRDEGGLDFELTSLAEAFVQGVPPVMTAHFEGGRPEFDLPFY- 3412  
QY 279 GEALARRVFRPQPDSSSLDAAFAADLLHAHFYESCPLYKFAHFTANQALILEAFAGCRV 338

Db 3413 --AFQRRYTWLHEEPLOEPPVEAW-----DAEWS----- 3440  
QY 339 HVNDFGIGQGMQWPAALLQALALRPGSPSFRITGVGPQPBETDALQOV-----GWLQAO 393  
Db 3441 -VERG-----DATAVSDDLSTDAELHTVLPALSSWRRR 3475  
QY 394 FAHTRVDFQYRGVLAATLADLEPMLQPEGEDEPNEEPEVIVANSVFEMHRLLAQPGAL 453  
Db 3476 VEHRLODMRR-----VEKKP-----PALDEVLGSGWLVVVRGLADDEV 3519  
QY 454 EKVIGTVRAVRPRIVVE-----QEAHNSG-----TFL--DR----- 485  
Db 3520 ARVVAAYTA-RGGEVSVVELPTRPDRRAYEAAVAGRCVSGVSPFLSMDDRHSEHSYVP 3578  
QY 486 --FTESLHYXTMDESLGSSSGGSPS--EVSSEA-AAAPAAAGT--QVWSEVY----- 533  
Db 3579 AGLASLVLAQALVDL-----GRAEGEPRMLVTRGAVVAGPSDAGVVIDPQAQWFGGRV 3635  
QY 534 LG-----ROICNVVACGAEETERETRTETLGGMRNRNLGNAGFETV 571  
Db 3636 LGLEHPELMGGLVDLPVGVDEVCRRFVGYVASAGFEDQVAVRGSGVVRRLVRA---VV 3692  
QY 572 HLGSNAYQASTLLALFAGDGYKYEKEGCLTIGWHTRPPLIATSAMRIAG 622  
Db 3693 DGGGCGWRPGRVTLV--TGGLG-----GLGAH-----TARMLVGG 3725

RESULT 11  
US-08-922-635-22  
; Sequence 22, Application US/08922635A  
; Patent No. 6033871  
; GENERAL INFORMATION:  
; APPLICANT: FILETZ, John E.  
; APPLICANT: IVANOV, Tina R.  
; TITLE OF INVENTION: DNA MOLECULES ENCODING IMIDALINE RECEPTIVE POLYPEPTIDES  
; FILE REFERENCE: Corrected Sequence Listing  
; Patent No. 6033871  
; CURRENT APPLICATION NUMBER: US/08/922.635A  
; CURRENT FILING DATE: 1997-09-03  
; EARLIER APPLICATION NUMBER: 08/650.766  
; EARLIER FILING DATE: 1996-05-20  
; EARLIER APPLICATION NUMBER: 60/012.600  
; EARLIER FILING DATE: 1996-03-01  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 1070  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-08-922-635-22

Query Match 3.8%; Score 120.5; DB 3; Length 1070;  
Best Local Similarity 20.0%; Pred. No. 0.2;  
Matches 105; Conservative 67; Mismatches 220; Indels 133; Gaps 20;

QY 121 QLNASTSTVT-----GSGG-----YFDLPVSDSSSIYALRPIPSAGATPADLS 168  
Db 566 QLRASLDLQKTVIAKTEGTGSPGSGFADGCPAERRASN--DGRPOQVPEALAPAVE 623  
QY 169 ADVSDPKRMRTGSGSTSSSSSSSSSSSGGARSSVEEAPVYAAAANATPALPVVVDTO 228  
Db 624 VPAPAPAAASASGPAKTPAPAEASTSALVPEETVPEADPAPAEAPADVPSSHLQATSE 683  
QY 229 E-----AGIRLVHALLACAFAVOOENLSAAEALVKQIPLLAAS 266  
Db 604 ENQIFSHLPACPSLRHVASLGSATIELFHSISIAEVEBELRLHMSSVYVYQFPGLEVT 743  
QY 267 QGAMRKVAAVFGAELARRVFRFPQPDSSSLDAAFAADLLHAHFYES--CPYLKFAHF--TA 324  
Db 744 ACVLISTKAAYF--VLHDLRRYFSEPLQDFWHQKNTYNNSPFHTISOCFVLKISDLSQV 801



QY	325	NQALE--AFAGRRVHYVDFGIKGMQMRPALLALRPGGPRFRLTGVGRPPDET	381
Db	802	NVGLFDQHFLLTGSTPMQVYTCITLRSTYLHCTLQILMYLVSS--LRRPSPRPDK	856
QY	382	DALQOVGWKLA-----QFAHTIRVDFOYRGIVAAATLADLERFMLOREGEEDRPNEBEVY	435
Db	857	DYSEEFENKTKTKMENVELHSSRVEFTYPS--EEIIGDLFTVAQKMAE--PEKAPALS	912
QY	436	AVNSVFEMHLLAQPALENVLTGTVAAVPRIVYVEQEAHNNSGFLDRFTESLHY--	492
Db	913	ILLYVAQFQVMPPPGCCRGPL--RKTYLLTSSEI--FLUDE--DCVHYPLP	959
QY	493	-----YST-----MFDLSLG-----GS	504
Db	960	EFAKEPQRDRYRIDGRIRVDRIDLVMGQYGFQALTLTVFDVQCHDLMSGVTLIDHGE	1011
QY	505	SGGSGSEVSSGAAA--APAAAGDQVWS-----EYVLGQI	538
Db	1020	VPGGPARASQREYQOMQVFPVPSSESEKILSLIARQEWALCGREL	1064

```

RESULT 12
US-09-428-517-3
? Sequence 3, Application US/09428517
? Patent No. 6251636
? GENERAL INFORMATION:
? APPLICANT: Betlach, Mary C.
? APPLICANT: Shah, Sanjay Krishnakant
? APPLICANT: McDaniell, Robert
? APPLICANT: Tang, Li
? TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
? FILE REFERENCE: 30062-20029, .00
? CURRENT APPLICATION NUMBER: US/09/428, 517
? CURRENT FILING DATE: 1999-10-28
? EARLIER APPLICATION NUMBER: 60/120, 254
? EARLIER FILING DATE: 1999-02-16
? EARLIER APPLICATION NUMBER: 60/106, 100
? EARLIER FILING DATE: 1998-10-29
? NUMBER OF SEQ ID NOS: 12
? SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 3
? LENGTH: 3816
? TYPE: PRT
? ORGANISM: Artificial Sequence
FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: Recombinant
? OTHER INFORMATION: Oleandolide PKS
? US-09-428-517-3

```

Query Match	3.7%	Score 119.5;	DB 4;	Length 3816;
Best Local Similarity	20.7%;	Pred. No. 1.6;		
Matches 145;	Conservative 71;	Mismatches 207;	Indels 279;	Gaps 35
QY	14	GGGGN--GSSDEKM-----MYSAAA-----GGEEDVDELLAL--GKYK 48		
		: : : :         : : :		
Db	2251	GGGGVAVSGLSAERVRTIMDTGGRVSVAAVNGPSTVYSGVQALDELLACERGVNA 2310		
QY	49	R-----ASDMAADVAKTLEOLEMANGMGCGVGAAPDDSFATHLATDVIHYNPT--DLS 99		
		:   : : : :   : : :   :   :   :		
Db	2311	RRVPVYASHSKOMQDLDELLEAL-----ADIPPOHSSVEFFSFVTADWIDTTALDAG 2364		
QY	100	SW-----VERMLSE-LNA-----PPPLPAPQ--LNA-STSSVYTG----- 133		
		: : : : :   : : : :   : : :   :		
Db	2365	YWFNTLRFTVRPQEAEEVGLVAGKMAFEVCECSHPALVPETLEOTLDALDONAAVLGSLRD 2424		
QY	134	-GGYEDLPSS-----YDSSSSITIALRP--IPS-----PAGATAPADLSA 169		
		:   : : : :   : : :   :   :   :		
Db	2425	EGGLDRILTLSLAFAVQGVPPVDMTHAFECMPTRTYDLPYPPROQHYYWKKRPAP----- 2479		
QY	170	DSVRDPKRRFTGSGSTSSSSSSSSSLGGGARSSVVEAAPVAAANATPALPVEVVVDQE 229		
		:   :   :   :   :   :   :   :   :		

```

Db 2480 -----GANTGDVAASGLTRAGPRLLGLGAVEMPDSDGLVL-TGQ 2515
QY 230 AGIRLVALLA-----CAEAQOENISAAEALVKQIPLIAASQGGAM 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2517 ISLR-THPWLDADHEVLGSLPLPGTAFAVELAQAADRAGYDVLDELTLBAPVLDPGSGIQ 2575
QY 272 RKVAAYTGCALARRVYFRFRQPOSSLLDAAFADLHAHYESCPILKFAHFTANALILEA 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2576 VRLALGPSEADGRSLQLHNSRPEA-----2600
QY 332 FAGGRVHVVDGFGKQOMQMPALLQLALRPGGPFSLRT---GVGPPQDETDAIQVG 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2601 -AGNHR-----WTRIASGFVY-PGCTGAARTEBAGVWPRPAGAEVALADSR 2645
QY 389 W-KIAQFAHTIRVDFQYGLVA------TLAD---LRFPMLOPGE--- 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2646 YARLVERGTYTGGPSFQ---GLHTAMRHGGDVYAEVALPECTPADGALHPALLDAVAQVG 2703
QY 426 -----EPPNE-----BEVIAVSVSEMHMLLAQPGAL 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2704 LGSFVEDPGVYUPEFLMSDVTYLHATGATSLRVRVSPAGPDVAL-----ALADPAGA 2755
QY 454 EKVGVTRAVRPRIVTVVEQEAHNHNGTFELDRTESLHYSTMFSLGSS-----G 506
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2756 P-VATYGALRLRTSAQ-----LARRGSAEHAEMRVEMVEBGSALADRCRGA 2803
QY 507 GGPSEVSSGAAAPAAAGTDVNSEVYLGRIQCINNVACEGAE 548
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2804 GCTTYEGERRAAEAGAAAGTMAV-----LGRVPAAVRTMGVD 2840

RESULT 13
US-08-976-255-11
: Sequence 11, Application US/08976255
: Patent No. 6136581
:
: GENERAL INFORMATION:
: APPLICANT: Jono, Keith E.
: APPLICANT: PLOWMAN, Gregory
: TITLE OF INVENTION: KINASE GENES AND USES
: NUMBER OF SEQUENCES: 53
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071-2066
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 MB
: MEDIUM TYPE: storage
: COMPUTER: IBM compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: FASTSEQ for Windows 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/976,255
: FILING DATE: No. 6136581ember 21, 1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/031,675
: FILING DATE: No. 6136581ember 22, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 229/182
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1384 amino acids
: type: amino acid

```

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-976-255-11

Query Match 3.7%; Score 119; DB 4; Length 1384;  
Best Local Similarity 21.5%; Pred. No. 0.38;  
Matches 120; Conservative 54; Mismatches 181; Indels 204; Gaps 28;

QY 116 LPPAPQLMASTSTVTGSGCYFDLP-----PSYDSSSYIALRPI 155  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 764 LAPAPCVTPBTWTASSCG--DHPQAEPLKATPAEGTTGPRLLPSPVPSQDEGA--PL 819  
QY 156 PSPAGATAPADLSDVDRPMRTGSSSTSSSSSSSLGGARRSVVEAPVAAAN 215  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 820 PSEE-ASAPD--ABDALPDSPTPATGE--VSAIKLASALGSSSPVEVA--PSEDED 872  
QY 216 ATPALPVVVDTQEAIRLVHALLACAEAVQENLSAAELVKOI----- 260  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 873 TAATSGIFDTSSDGL-----QARRPDVVPAPRSIQKQVGPDSLSDIPSSA 922  
QY 261 -----PLAASOGGAMKV-AAIFGEALARKVFRFPQPSLIDA-----AFA 303  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 923 SDGGYEVFSPSATGSPGQPRALDSYDTEHY-----ESPEVLKEAGECEPQAF 974  
QY 304 DLH-----AHFYESCPYLKFAHTANQALILEAFAG---C--RRVHYD 342  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 975 ELASEGEGPPELTSLSLGSLNKNPTROSAITFSDLEAEKATSGPEKKGGRAPGE 1034  
QY 343 FGKIQGMQWALLQALALRPG-----PSEFRLTGVG-----PPD 379  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 1035 LGLPSTGQ-PS-DQVCLRPGEVSGEAGSGRGEVLPPLQEGSSPEPSTGSLVPEPP 1091  
QY 380 ETDLDQV-----GKMLQFAHTIRVDFOYKGLVAATLADLEPMLQPEGEDPNEPEVY 435  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 1092 EPQGPAPKRPSPSSCSQFPL-----LTPVPLRSNGSSEFGGP-- 1131  
QY 436 AVNSVFEMHRLAOPGAEKVIQ-----TVRAVPRIVTVE--QEAHNSGFLDERFT 487  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 1132 -----GLSGPAPQKRMGCPGTAPRRLALPGLPALLEGPEEEDSDSDSD 1182  
QY 488 ESLHYSTMEDSLGSSGGSPSEVS--SGAAAPAAAGTDQ-----VMSEVY 533  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 1183 BELCYSYQ-----PSESEDEEAPAVPVVAESASANRLSLKMPSLSETF 1231  
QY 534 LGRQICNVACEGAERTER 552  
DB 1232 -----CEDLERKK 1240

## RESULT 14

US-08-804-227C-14  
; Sequence 14, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4545 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-804-227C-14

Query Match 3.7%; Score 119; DB 2; Length 4545;  
Best Local Similarity 21.7%; Pred. No. 2.3;  
Matches 143; Conservative 79; Mismatches 223; Indels 214; Gaps 35;

QY 9 GSGSGG---GGGM---GSSSEDKM-----MVSAAAGEGEE-----VDELLAA 43  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 1731 GGLMGLPYVGCGMMSVGAASEVYGVGEGLEWVSVAAVNGPRSVLSGDVGLSEVYAS 1790  
QY 44 L---GYKVRASDMA-----DYAOKLEOLEAMKM-----GCVGAGA 76  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 1791 LMGDVGEGRILDVSHGFHSLMEVLEGFRCGVESLEGRVPRGVVSVSGVSVSGE 1850  
QY 77 APDSEFATHLATDVRVHY-----NPTDLSWVE-----SMUSELMAPPPLPAPQ 121  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 1851 LGDPGYVRRARREAVRFADGCVYRGGLGVLVEGPRGVLTGAGELGAGDDVYVPA 1910  
QY 122 L-----NASTSTV--TGS-GGYFDLPSPVSSSYIALRPIPSP 158  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 1911 MRGRAREVEFEALATVFTADGADLTALHTGSTRIDL-PTYPQORDRYWLDPVTA 1969  
QY 159 AGATAPADLSDVDRPMRTGSSSTSSSSSSSLGGARRSVVEAPVAAANATP 218  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 1970 VTGVEPAGSPADA-----RATERGRSTTA-----GIRYNA-WQPAVVDGRNGP 2013  
QY 219 ALPVVV---DTEAGIRLVHALLACAEAVQOE-NLSAAELVKOIPLLAASOGAMRK 273  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 2014 AGHULLAPDEDTADSGL-----APATARELAVRGAEVHVAVPVCGREAGDGL 2064  
QY 274 VAAYFGEALARKVFRFP--QPDSSLLDAAFADLHAHFYESCPYLKFAHTANQALILEAF 332  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 2065 RAAGDGAARSTVLMLAPAEPDAA--DAV-----ALVQALGEA- 2100  
QY 333 AGCRVHVVDGFIQGMQWALLQALALRPGSPFRITGVGPQPDETDALQGVKMLA 392  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 2101 -----VEAPLWITTRERAAVPRDETPS-----VGAGQL-----MGLG 2133  
QY 393 QFAHTIRVDFOYRGLV---AATLADLEPFM--LOPEGEEDPNEPEVYIAVNSFEMHRL 446  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 2134 QVA-ALEIGRRMGGLADLPASPAVLRTFYGALLAGGENOFAPRPSVHRRVVP----- 2188  
QY 447 LAOPGAEKVLGTVRAVPRIVTVEQEAHNSGFTL--DRETESLHYSTMEDSLGEG 503  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 2189 -----APVPVPAASARTVTTATATAGEDARDTSDVVVPDDRMSSG-----TVLITG 2235  
QY 504 SSGGSPSEVS-----SGAA-----AAPAAGTDQWSEV-YLGRQICNVACEGAER 549  
| | | | : : : : | | | | : : : : | | | | : : : :  
DB 2236 GTGALGAQVARRLARSGAARLLLVGRGAAGPVGELVEELTALGSEVA-VEACDVADR 2293

## RESULT 15

US-08-804-227C-8  
; Sequence 8, Application US/08804227C  
; Patent No. 5876991

```

GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kubstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCE ADDRESSES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4550 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-804-227C-8

```

```

Query Match 3.7%; Score 119; DB 2; Length 4550;
Best Local Similarity 21.7%; Pred. No. 2.3; Mismatches 223; Indels 214; Gaps 35;
Matches 143; Conservative 79;

QY 9 GGGGGG---GGGM---GSSSEDKM-----MWSAAGEGEE-----VDELLAA 43
DB 1736 GGLMGLPVGGMWSVGAESVRYVEGIGEVSVAAVNGPVSVALSGDVGLSESVAS 1795
QY 44 L---GYKVRASDMA-----DVAOKLEOLEMANGM-----GGVGAGA 76
DB 1796 LMGDGYECRLDVSHGPHSVLMEPVIGEFGVSELEFGHVRPQVYVSGVSGVGSSE 1855
QY 77 APDDSFATILATITVH-----NPTDLSWYE-----SKLSELNAPPPPLPPAPQ 121
DB 1856 LGDPGYVWVRHAREAVRFADGVVGRGLGVTLYEVGPHGYLTGMAGCICGAGDDVVVPPA 1915
QY 122 L-----NASTSSTV---TGS-GGYFDLPSSVDSSSITVALRPIPS 158
DB 1916 MRGRAREVEFEALATVFTTRDGLDNTALHTGSTGRIDL-PPYPPQRRDYWLDPVRTA 1974
QY 159 AGATAPADLSADSVRDPKRMRTGSSSTSSSSSSSSSISGGARSVVEAAPPVAAAANATP 218
DB 1975 VTGVEPAGSPADA-----RATERGRSTTA-----GIRYRVA-WQPAVVDGRGNPGP 2018
QY 219 ALPVVVV---DTQEGIRLVHALLACAEAVQOE-NISAAEALVKQIPLLAASOGGAMRK 273
DB 2019 AGHVLLAPDEPDADSGL-----APAIARELAVGAEEHTVAVPGVTGREAGDGL 2069
QY 274 VAAVFEALARRVFRFRP-OPDSSILDAFAADLLHAHFYESCPLYKFAHFTANQALIEAF 332
DB 2070 RAAGDGAARSTRVLMALPAEPDAA---DAY-----ALVOALGEA- 2105
QY 333 AGCRHVHVDFGIKQGMQWMPALLQALALRPGPPSFRLTGVGPQDPDETALQOVGWKLA 392
DB 2106 -----VPEAPLMTTTRAAAVRPDETPS-----VGGAQL-----WGLG 2138

```

```

QY 393 QFAHTIRVDFOYRGLV-----AATLADLEPEM--LOPEGEEDPNEEPEVIAVNSVFEMHRL 446
DB 2139 QVA-ALELGRRMGGLADLPSSASPAVLRTVYGALLAGGEQFAVRSGVARRVVP----- 2193
QY 447 LAOPGALKYLGTVRAVRPRIVTVVQEOANHNSGTFP---DRFTESLHYYSYMFDSIEGG 503
DB 2194 -----APVPVPASARTVTTAPATAVGEDARNDTSDVVPDPDRWSSG-----TVLITG 2240
QY 504 SSGGGRSEVS-----SGAA-----AAPAAAGTDQYMSFV-YLGRQICNVVACEGAER 549
DB 2241 GTGALGAQVARRLARSQAARLLVGRGAAGPGVGLVEBELTALGSEVA-VEACDVAADR 2298

```

Search completed: January 22, 2002, 16:24:19  
Job time: 39 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2002, 23:03:40 ; Search time 3021.93 Seconds  
(without alignments)  
11600.701 Million cell updates/sec

Title: US-09-485-529-14

Perfect score: 2125  
Sequence: 1 atagagagcgagcgtagctc.....ttgagtgcagcgactccg 2125

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_ov: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sy: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_da: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_om: \*  
20: em\_or: \*  
21: em\_ov: \*  
22: em\_pat: \*  
23: em\_ph: \*  
24: em\_pl: \*  
25: em\_ro: \*  
26: em\_sts: \*  
27: em\_sy: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_htgo\_hum: \*  
31: em\_htgo\_inv: \*  
32: em\_htgo\_rnd: \*  
33: em\_htg\_hum: \*  
34: em\_htg\_inv: \*  
35: em\_htg\_rnd: \*  
36: em\_htg\_other: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2125	100.0	2125	6	AX005805	AX005805 Sequence
2	1956.2	92.1	2709	6	AX005794	AX005794 Sequence
3	1870.4	88.0	1872	8	TA8242531	AJ242531 Triticum
4	1502.4	70.7	1768	6	AX005804	AX005804 Sequence
5	1379.8	64.9	1746	6	AX005848	AX005848 Sequence
6	1341.4	63.1	2500	8	AB030956	AB030956 Oryza sat
7	1341.4	63.1	122497	8	AC087977	AC087977 Oryza sat
8	1317	62.0	2225	6	AX005806	AX005806 Sequence
9	1309.4	61.6	1890	8	ZMA242530	AJ242530 Zea mays
10	120.2	29.2	800	8	AF377646	AF377646 Trispacum
11	617	29.0	800	8	AF377621	AF377621 Zea mays
12	617	29.0	800	8	AF377624	AF377624 Zea mays
13	617	29.0	800	8	AF377627	AF377627 Zea mays
14	617	29.0	800	8	AF377633	AF377633 Zea mays
15	617	29.0	800	8	AF377645	AF377645 Zea mays
16	615.4	29.0	800	8	AF377626	AF377626 Zea mays
17	615.4	29.0	800	8	AF377631	AF377631 Zea mays
18	615.4	29.0	800	8	AF377634	AF377634 Zea mays
19	615.4	29.0	800	8	AF377635	AF377635 Zea mays
20	615.4	29.0	800	8	AF377636	AF377636 Zea mays
21	615.4	29.0	800	8	AF377639	AF377639 Zea mays
22	615.4	29.0	800	8	AF377640	AF377640 Zea mays
23	615.4	29.0	800	8	AF377642	AF377642 Zea mays
24	615.4	29.0	800	8	AF377643	AF377643 Zea mays
25	613.8	28.9	800	8	AF377622	AF377622 Zea mays
26	613.8	28.9	800	8	AF377623	AF377623 Zea mays
27	613.8	28.9	800	8	AF377625	AF377625 Zea mays
28	613.8	28.9	800	8	AF377628	AF377628 Zea mays
29	613.8	28.9	800	8	AF377629	AF377629 Zea mays
30	613.8	28.9	800	8	AF377630	AF377630 Zea mays
31	613.8	28.9	800	8	AF377632	AF377632 Zea mays
32	613.8	28.9	800	8	AF377637	AF377637 Zea mays
33	613.8	28.9	800	8	AF377638	AF377638 Zea mays
34	613.8	28.9	800	8	AF377641	AF377641 Zea mays
35	613.8	28.9	800	8	AF377644	AF377644 Zea mays
36	420.4	15.8	770	6	AX005803	AX005803 Sequence
37	382	18.0	132699	6	AC006917	AC006917 Genomic s
38	381.8	18.0	1964	6	ATY15193	A64697 Sequence 1
39	381.8	18.0	1964	6	ATY15193	Y15193 Arabidopsis
40	380.2	17.9	1951	8	ATRG42	Y11337 A. thaliana
41	380	17.9	1779	6	AX081276	AX081276 Sequence
42	378.4	17.8	1779	6	AX081278	AX081278 Sequence
43	367	16.3	416	6	AX005809	AX005809 Sequence
44	359.8	16.9	26604	8	ATK3M16	AL391150 Arabidops
45	359	16.9	436	6	AX005863	AX005863 Sequence

#### ALIGNMENTS

RESULT 1  
LOCUS AX005805 2125 bp DNA PAT 24-AUG-2000  
DEFINITION Sequence 14 from Patent WO9909174.  
ACCESSION AX005805  
VERSION AX005805.1 GI:99288602

#### KEYWORDS

SOURCE

#### ORGANISM

bread wheat.  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooidae; Triticeae; Triticum.

#### REFERENCE

1 (bases 1 to 2125)  
Harberd, N.P. and Peng, J.

#### AUTHORS

Genetic control of plant growth and development  
Patent: WO 9909174-A 14 25-FEB-1999;

#### TITLE

HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)

#### FEATURES

source

1..2125  
/organism="Triticum aestivum"

BASE COUNT	325	a	768	c	723	g	309	t
ORIGIN	/db_xref="taxon:4565"							

Query Match	100.0%;	Score 2125;	DB 6;	Length 2125;
Best Local Similarity	100.0%;	Pred. No. 3.5e-213;		
Matches 2125;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible][illegible]

Db	2041	GAATGCAAGCAGCCGGCTACGATCGATCGGCGCTACGGGTTCCCTCCTGCGCTGA	2100
Qy	2101	agaagtgagatgacgacgacctcg 2125       	
Db	2101	AGAGGTGATGACGACGACACTCCG 2125	
RESULT	2		
AX005794			
LOCUS	AX005794	2709 bp DNA	PAT 24-AUG-2000
DEFINITION	Sequence 3 from Patent WO9509174.		
ACCESSION	AX005794		
VERSION	AX005794.1	GI:9928799	
KEYWORDS			
SOURCE			
ORGANISM			
	bread wheat. Triticum aestivum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Poidea; Triticeae; Triticum. 1 (bases 1 to 2709)		
AUTHORS	Harberd,N.P. and Peng,J.		
TITLE	Genetic control of plant growth and development		
JOURNAL	Patent: WO 9509174-A 3 25-FEB-1999; HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)		
FEATURES			
source	1..2709 Location/Qualifiers		
misc_feature	/organism="Triticum aestivum" /db_xref="taxon:4565"		
misc_feature	6 /note="n is any nucleotide"		
misc_feature	11 /note="n is any nucleotide"		
misc_feature	46 /note="n is any nucleotide"		
misc_feature	85 /note="n is any nucleotide"		
misc_feature	107 /note="n is any nucleotide"		
misc_feature	407 /note="n is any nucleotide"		
misc_feature	417 /note="n is any nucleotide"		
misc_feature	430 /note="n is any nucleotide"		
misc_feature	436 /note="n is any nucleotide"		
misc_feature	448 /note="n is any nucleotide"		
misc_feature	452 /note="n is any nucleotide"		
misc_feature	490 /note="n is any nucleotide"		
misc_feature	498 /note="n is any nucleotide"		
misc_feature	556 /note="n is any nucleotide"		
misc_feature	562 /note="n is any nucleotide"		
misc_feature	727 /note="n is any nucleotide"		
misc_feature	769 /note="n is any nucleotide"		
misc_feature	1444 /note="n is any nucleotide"		
misc_feature	1450 /note="n is any nucleotide"		
misc_feature	1556 /note="n is any nucleotide"		
misc_feature	1742 /note="n is any nucleotide"		
misc_feature	1808 /note="n is any nucleotide"		

	misc_feature	1900	/note="n is any nucleotide"	
	misc_feature	1940	/note="n is any nucleotide"	
	misc_feature	1982	/note="n is any nucleotide"	
	misc_feature	2438	/note="n is any nucleotide"	
	misc_feature	2604	/note="n is any nucleotide"	
	misc_feature	2638	/note="n is any nucleotide"	
	misc_feature	2681	/note="n is any nucleotide"	
	misc_feature	2689	/note="n is any nucleotide"	
	misc_feature	2706	/note="n is any nucleotide"	
BASE COUNT	473 a	900 c	823 g	482 t
ORIGIN			31 others	
Query Match	92.1%;	Score 1956.2;	DB 6;	Length 2709;
Best Local Similarity	96.8%;	Pred. No. 1.2e-195;		
Matches 2061; Conservative	0;	Mismatches 53;	Indels 15;	Gaps
OY	1	atagagaggcggaggtgacgtccgcggaatcaagaagcgaggattaccagagacgcggaggagac	60	
Db	135	ATAGAGAGGCGAAGTACTCCGGATCATGAACGGGACTACCAGAGCCTCGGAGGAGAC	194	
OY	61	ggcgcgcgcgctgtgcgatggcttcgtctcgcagagacaagatgatgtgtcgcgcgcgcgc	120	
Db	195	GGCGCGGGGGTGCGGCATGGGTTCGTCGCCAGACAAGATGATGATGTGTGCGCGCGGCG	254	
OY	121	ggggagggggaggagggttgacgactgtctgcgcgcgtcgcgttacaaggtygcgcctcc	180	
Db	255	GGGGAGGGGGAGAGAGTAGTGCACAGCTGCTGGCGGCGCTCGGGTTAAAGGTGCGCCTTC	314	
OY	181	gacataggcggaagtgtggccagaagactltgaaagagctcagagatgcatatgggaatggcggc	240	
Db	315	GACATAGGCGGAGGTGAGCGGAGAGAGCTGGAGGCACTCGAGATGGCAATGGGATGGCGGC	374	
OY	241	gttgggc	300	
Db	375	GTGGGCGCTGGGGCGCGCCCTCAGCACAGATTNGCCACCC -GCNMGCGCGGACACMNTG	433	
OY	301	cactacaaccccacccagacctgtctctcttgtgtgcagagacatgtctgtcggagctcaagcgc	360	
Db	434	CANTACAACCCCCACAGNCAGMTGTCGTTGGTGGAGAGCATCTGTGAGACTTAANGAG	493	
OY	361	cgc	420	
Db	494	CGCGNCGCGCCCTCCC CGCCCGCCCGCAGCTCAACG -----CCTCAACGCTCAACG	544	
OY	421	ggcagc	480	
Db	545	GGCAGCGGGGGATTACTTNATATCTCCGCGCCTTAGTGCAGTCTCCAGCAGCATCTAACGGG	604	
OY	481	ctgc	540	
Db	605	CTGCGGCGGATCCCTCCCGCGCGCGCGGAGAGGGCGCGCGGACCTGTCCCGCAGCTCC	664	
OY	541	gtgc	600	
Db	665	GTGCGGGATCCCAAACGGAGTGGCACTGGCGGGAGAGCAGCACTCGTCGTATCTCTCTA	724	
OY	601	tctgtctctctctgt	660	
Db	725	TANMTGTCCTCTGTGGTGGGGCGCCAGAGACTGTGTGTGTGTAAGCAGCCCGCGGTCGG	784	
OY	661	gccgc	720	
Db	785	GCCGCGGCGCAAGCACAGCCCGCTGCGCGTCTGCTGTGTGTGAACAGCAGAGAGCGGG	844	

[illegible][illegible][illegible]

RESULT	3
LOCUS	TAE242531
DEFINITION	TAE242531 1872 bp DNA PLN 28-JUL-1999
ACCESSION	Triticum aestivum rht-D1a gene for gibberellin response modulator.
VERSION	AJ242531
KEYWORDS	AJ242531.1 GI:5640156
SOURCE	gibberellin response modulator; rht-D1a gene. bread wheat.
ORGANISM	Triticum aestivum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
REFERENCE	Pooideae; Triticeae; Triticum. 1 (bases 1 to 1872)
AUTHORS	Peng,J., Richards,D.E., Hartley,N.M., Murphy,G.P., Devos,K.M., Flintham,J.E., Beales,J., Fish,L.J., Worland,A.J., Pelica,F., Suchanek,D., Christou,P., Snape,J.W., Gale,M.D. and Harberd,N.P. 'Green revolution' genes encode mutant gibberellin response modulators Nature 400 (6741), 256-261 (1999)
JOURNAL	99347734
MEDLINE	2 (bases 1 to 1872)
REFERENCE	Harberd,N.P., Peng,J. and Richards,D.E. Green revolution genes encode mutant gibberellin response modulators Unpublished 3 (bases 1 to 1872) Richards,D.E. Direct Submission Submitted (25-MAY-1999) Richards D.E., Molecular Genetics, John Innes Centre, Colney Lane, NR4 7UJ, UNITED KINGDOM
JOURNAL	Title
AUTHORS	
FEATURES	Location/Qualifiers 1..1872 /organism="Triticum aestivum" /db_xref="taxon:4565" 1..1872 /gene="rht-D1a" 1..1872 /gene="rht-d1a" /note="GAI ortholog" /codon_start=1 /product="gibberellin response modulator" /protein_id="CAB51555.1" /db_xref="GI:5640157" /translation="MKREYDAGGGGGCGSSSEDKMMVSAAGEGEVEDELLAAL CYKRSADMDVNAQLEOLEAMGMCGACAAAPDSFATLADTVHYHNPDISMWA ESMLSELNAPPPLPPAQUNASTSYVTGSGETFDLPVSVDSSSTIATLRPISPAQ ATAPADVSADVROPKMRRTGSSSTSSSSSSSLGGARSSVVEAAPPVAAAANAFT ALPVAVVDTDEGRITRLVALACAEAVOENLSAAELVKQIPILLAASOGAMRKVAA
gene	
CDS	





[illegible][illegible]



nuclear-localized protein capable of gene activation at  
transcriptional level  
Gene 245 (1), 21-29 (2000)  
JOURNAL  
MEDLINE  
20179680  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Sano, H., Ogawa, M. and Kusano, T.  
Direct Submission  
Submitted (09-AUG-1999) to the DDBJ/EMBL/GenBank databases.  
Mikihito Ogawa, Nara Institute of Science and Technology, Plant  
Molecular Breeding, Takayama 8916-5, Ikoma, Nara 630-0101, Japan  
(E-mail:m-ogawa@ds.aisi-nara.ac.jp, Tel:+81-743-72-5652,  
Fax:+81-743-72-5659)  
FEATURES  
source  
1. .2500  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/dev\_stage="seedling"  
216. .2093  
/gene="Osga1"  
216. .2093  
/gene="Osga1"  
/note="GAI homolog"  
/product="Osga1"  
/protein\_id="BAA90749.1"  
/db\_xref="GI:6970472"  
/translation="MKREYDAGSGSGSSADMGKDYKMACAGEEDVDLLA  
LGIVRSSDADVAQKLEQLEMAKMGVSAPGAADGPFVSLATDVHYHPSLSM  
VESMLSELNAPLPPIPPAPPARHASTSVTGGSGFELPAADSSSTVALRP  
SLPVATADPSPADSDARTKMRGTGSSSSSSSLGASGRSVLEAPATOG  
AAANAVPVPVVVDTOEAGIRLHVALLACAEAVOEFAAEALVKQIPFLASOG  
AMRVAVFGEALRAVYRFRPADSTLLDAFADLLAHFESCPYLFKFAHTANQAI  
LEAFAGCHRVHVFEGIKQGMWPALLQALALRGGPSPFLTVGPPQEPETALQO  
VGMLEAFHRTIRVDPOYRGIVATLADIDPEPMIOPGEADANEPEPVANSFELH  
RLAQPGALEKVLCTVHAHRRIVTVVBOENHNSGFLDRFTSLHSTYMPFSLG  
GSSGQALSPPAAGGGGTDQVMSVYVLRQICNVVACEGERHERHETLQOMNRIG  
RAGEPVHLGNAVKQASTLLALPAGDGVREBEKGLTGMWTRPLIANSARVAA  
A"  
BASE COUNT 413 a 861 c 771 g 455 t  
ORIGIN

Query Match 63.1%; Score 1341.4; DB 8; Length 2500;  
Best Local Similarity 84.6%; Pred. No. 1.3e-131;  
Matches 1640; Conservative 0; Mismatches 251; Indels 48; Gaps 10;

QY 2 tagaagagcgaggtatgcgcgagatcatgaagcggagatcaccagagccggagagagcg 61  
DB 190 TAGGTAGGTTGGGGGAGCGAGATCATGAACCGGAGTACCAAGACCGCGGGAGACA 249  
QY 62 gcgagcgagctgagcgcc-----atgggctcgtccgagagacaagaatgagtgcgagc 115  
DB 280 GCGGCGCGCGGAGCGACGCCGATATGGGGTCTGCAAGGACACAGGTGATGC---CGGGG 306  
QY 116 cggcggggagggagggagagtgagcagcgtgctggcgagcgtcggagtcacaagtgccg 175  
DB 307 CGCGCGGGGAGGAGGAGACGTGACGAGCTGCTGGCGCGCTCGGATCAACAAGTGGCGGT 366  
QY 176 cctccgaatagtcggaagctgagcgaagctgagcagctcgaagatgagcgaatggagatgg 235  
DB 367 CGTCCGATGCGCCACGCTCGCGGAGAGCTGAGACACTGAGATGGCCATGGGAGATGG 426  
QY 236 gcgagctggagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 295  
DB 427 GCGGCGGTAGCGCCCGCGCGCGCGGATGAGAGGGTTGCTGTCGACACTGGCCACAGGACA 486  
QY 296 ccgtgactacaacccacacgacgtgctctggtgctgagagcagtgtctgtcgaagctca 355  
DB 487 CCGTCACTACAAACCCCTCGGACCTCTCTCGGTGAGAGCATGCTTCCGAGCTCA 546  
QY 356 acgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 409  
||||||| ||| || ||||| || ||| || ||||| ||

DB 547 ACGCGCCGTGCCCCCTATGCCGCGAGCGCGCGGTGCCCCCATGCTTCCACCTGCT 606  
QY 410 caaccgtacagggagcg-----gcggctactcatctcccgccctcgctgactct 463  
DB 607 CCACGTGATCCGGCGGCGGTGTAGCGGCTTCTTGAACTCCACACCGCTGCGGACTGCT 666  
QY 464 ccagcagatcatcgcgctgcgccgcatccctcccgcgcgagcgagcgagcgagcgagcg 523  
DB 667 CGAGTAGACACTACGCCCTCAGCGCGGATCTCTTACCGGTGTGTGCGACG---CTGACC 723  
QY 524 accgtccgagcagctcgtgctgagatcccaagcgagatgcgactggtggtgagcagacct 583  
DB 724 CGTCCGCTGTGACTCGCGGAGGAGACACACCGGATCGGCACTGGCGCGGCGCACACGT 783  
QY 584 cgtgcatccctccctccctcgctgctct---cggtggggggcgagggagctctgtgtgg 640  
DB 784 GTCCTCTCATGCTGCTGCTTCTTCTTGCGGGGTGGGCTCGCGGGGCTTGTGTGTG 843  
QY 641 agcgtgcccgcgcg-----tcgagcgcgcgcccaagcgagcgcgcgcgtgctgctg 694  
DB 844 AGGCTGTCGCCCGCGGACGCGAAGGGCGCGCGCGCGGCGAATGCGCGCGGCTTGGGTTG 903  
QY 695 tctgtgtcgaacgcagagagcgagatcgtgtgtgcaagcgtctgtgtgcgtgcgcg 754  
DB 904 TGGTGTGTGACACGAGAGGCTGGGATCCGCTGTGCACCGCTGTGTGGCTGCGCGG 963  
QY 755 aggcgttcagagagagaaacctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 814  
DB 964 AGCGCTGTCAGCAGAGAACTTGGCGCGCGCGGAGCGCTGTCAAGCAGATCCCAAGC 1023  
QY 815 tggcgcgctcccaagggcgagcagatgacgaaggtgcgcgcgcgcgcgcgcgcgcgcgc 874  
DB 1024 TGGCGCGCTCCAGGGCGGCGCATGCCAAGGTGCTGACTTCTGCGGCGGCGGCTTGC 1083  
QY 875 cccgcgcgctctcgcgtctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 934  
DB 1084 CCGCGCGGTGATCGGCTTCCGCC---CGGACAGACACCTCTTCAAGCGCGGCTTGC 1140  
QY 935 ccgacctctcagcgagcactctcagatctcgtccctaactcaagttcgcgacttca 994  
DB 1141 CGGACTTCTGACGCCGCACCTTCTAGAGTCTGCGCTACTCAAGTTCGCGCACTTCA 1200  
QY 995 ccgcaacacagcactctctgagagcgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1054  
DB 1201 CCGCAATTCAGCCACTCTCGAGGCTTTCGCGCGCTGCGCGGCTCCACGCTCGACT 1260  
QY 1055 tccgataaagagagagtgagtgagtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1114  
DB 1261 TCGGATCAAGGAGGGATGCAATGGCCAGCTCTCTCCAGGCTCTCGCCTTCTGTCCG 1320  
QY 1115 gcgccctccctcgtctcgcctcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1174  
DB 1321 GCGGCGCCCAACGCTTCGCGCTCACCGGCGTGGCGCCCGCGAGCGGACGAGACGAGC 1380  
QY 1175 ccctgcagcaggttggtctggaagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1234  
DB 1381 CTTGACACAGGTGGTGGAACTTGGCCAGTTGCGCACACCAATTTGGCGTGGCTTTC 1440  
QY 1235 agtaccgagcgtctgctccgc 1294  
DB 1441 AGTACCGGAGACTGTCGCGCGCTGCGCTGCGACTTGGAGCGCTTCAATGCTGACGGCG 1500  
QY 1295 aggcgaggaagaccgagacgagagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1354  
DB 1501 AGGCGAAGCGGAGCGGACGAGAGAGAGCTGAGAGTATGCGCGTCAACTCGGGTTCGAGC 1560  
QY 1355 tgcacgcgctgctgc 1414  
DB 1561 TGCACGCGCTGCTCGCAGCGCGCGCGCTGGAAGAGTCTCTGGGACAGGCTCACGCGG 1620  
QY 1415 tgcgcgcagagatgctacgcggtgtgtgagcagagagcgaatcaaatcgcgcgcgcgcgcgc 1474  
DB 1621 TCGGCGCAAGGATGCTCACCGGTGTGAGACAGGAGCCAAACCAATCTCGGCTCATTTCC 1680  
||||||| ||| || ||||| || ||| || ||||| ||



[illegible][illegible]





[illegible][illegible]





[illegible]

OY	963	tgcgcaacttaacgccaccagccatctcggaaggcgttcggcggctgcgcgcgcgtgc	1042
Dd	481	TGCGCCACTTCACTAGTCGAACACGAGGCATCTTCGAGAGCTTTGGCGGGCCGCCGCGCTCC	540
OY	1043	acgttgctgactctggataaacagcaggatgcagtggccgcacttctcaagcccctcg	1102
Dd	541	AcGTtGCCGaCtTTCGCaTfCAAGcAGGGGaTGCAtGTGcCcGCTTtTCCAGGcCCTTCG	600
OY	1103	cctctgcgtccggcggccccctccctcgttcgcgcctcaaccggcgctgcgcccccgcagccgg	1162
Dd	601	CcCTtCGGcCtTGgGgGgCCCCCATGtTTCGcGtTACcGgGgGTGGtTCGcCGcAGccCGG	660
OY	1163	acgagaaccgaagccctcgcagcagatggcctygaagctgcgccagttgcgcacacatcc	1222
Dd	661	ACGAaACCgAGcCCCTCGcAGAGGTGGGGCTGGAaAATTGCCAGTTCGcGcACAACATCC	720
OY	1223	gcgtgactttcacafraaccgcggcgcctcgtcgcgcgcacagctgcgcgacctgatgacgttca	1282
Dd	721	GcGTtGCACTTTCAGTAfCACGTGcCTGfTGGcGCGCACcCTGcGTGACCTGGAAGcCCTTCA	780
OY	1283	tgcctgcagccgaaaggcca	1301
Dd	781	TGCTCAACCGAGGGCCA	799
RESULT	11		
AF377621			
LOCUS	AF377621	800 bp DNA	PLN 01-AUG-2001
DEFINITION	Zea mays subsp. mays isolate GUAL14 gibberellin response modulator gene, partial cds.		
ACCESSION	AF377621		
VERSION	AF377621.1	GI:14318114	
KEYWORDS	.		
SOURCE	maize.		
ORGANISM	Zea mays subsp. mays		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC		
	clade; Panicoidae; Andropogonae; Zea.		
	1 (bases 1 to 800)		
REFERENCE	Tenallion,M.T., Sawkins,M.C., Long,A.D., Gaut,R.L., Doebley,J.F.		
AUTHORS	and Gaut,B.S.		
TITLE	Patterns of DNA sequence polymorphism along chromosome 1 of maize		
JOURNAL	(Zea mays ssp. mays L.)		
PUBMED	Proc. Natl. Acad. Sci. U.S.A. 98 (16), 9161-9166 (2001)		
REFERENCE	11470895		
AUTHORS	2 (bases 1 to 800)		
TITLE	'Tenallion,M.T., Sawkins,M.C., Gaut,R.L., Long,A.D., Doebley,J.F.		
JOURNAL	and Gaut,B.S.		
	Direct Submission		
	Submitted (07-MAY-2001) Ecology and Evolutionary Biology,		
	University of California, Irvine, 321 Steinhaus, Irvine, CA		
	92697-2525, USA		
FEATURES			
source	Location/Oualifiers		
	1..800		
	/organism="Zea mays subsp. mays"		
	/cultivar="Serrano"		
	/isolate="GUA14"		
	/sub_species="mays"		
	/db_xref="taxon:4578"		
	/chromosome="1"		
	/country="Guatemala"		
	<1..>800		
	/product="gibberellin response modulator"		
	<1..>800		
	/note="v8"		
	/codon_start=3		
	/product="gibberellin response modulator"		
	/protein_id="AAK59901.1"		
	/db_xref="GI:14318115"		
	/translation="PSADPSTDSAREKKRKRRTGGCTSSSSSSSSMDGRTSSVYE		
	AAPPATQASAANGPAVVVVVDIQEGIGLVALLAACAAVOENFSAEALVKDIP		
	MLSPSGGARAKRYVAAYGEALAIRRVYFRPRPPSSILDAFAFDLLAHFECSPIYKF		
mRNA			
CDS			



[illegible][illegible]

RESULT	14
LOCUS	AF377633
DEFINITION	AF377633 800 bp DNA PLN 01-AUG-2001
ACCESSION	AF377633
VERSION	AF377633.1
KEYWORDS	maize.
ORGANISM	Zea mays subsp. mays
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
AUTHORS	1 (bases 1 to 800)
TITLE	Tenallion, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F. and Gaut, B.S.
JOURNAL	Patterns of DNA sequence polymorphism along chromosome 1 of maize
PUBMED	(Zea mays ssp. mays L.)
REFERENCE	Proc. Natl. Acad. Sci. U.S.A. 98 (16), 9161-9166 (2001)
AUTHORS	2 (bases 1 to 800)
TITLE	Tenallion, M.I., Sawkins, M.C., Gaut, R.L., Long, A.D., Doebley, J.F. and Gaut, B.S.
JOURNAL	Direct Submission
PUBMED	Submitted (07-MAY-2001) Ecology and Evolutionary Biology,
REFERENCE	University of California, Irvine, 321 Steinhans, Irvine, CA
AUTHORS	92697-2525, USA
FEATURES	Location/Qualifiers
Source	1..800
	/organism="Zea mays subsp. mays"
	/cultivar="Bolita"
	/isolate="OAX68"
	/sub_species="mays"
	/db_xref="taxon:4578"
	/chromosome="1"
	/country="Mexico: southern"
	<1..>800
mrna	/product="gibberellin response modulator"
	<1..>800
	/note="d8"
	/codon_start=3
	/product="gibberellin response modulator"
	/protein_id="AAK59913.1"
	/db_xref="GI:14318139"
	/translation="PSADPSTDSARERKRMFTGGGTSSTSSSSSSMDGCTRSSVYE
	AAPATASAANGPAAVPPVVVDVQEGIRLVALLCAEAVQDENSAAALVKQILE
	MSGGGAMKVAAYFGEALRRVYRFRPPDSDLDAFADILHAFHFEPCYKLFKE
	ALHFANQALILEAFGCRRVHVVDGIRIKOGMOMPALDALALRPGSPFSFRTGVGPPQ
	PDEFDALDOYGWKLAFQFAHTRIVDFOYRGVLVAATLADLEPMLQPED"
BASE COUNT	105 a 305 c 262 g 128 t
ORIGIN	
Query Match.	29.0%; Score 617; DB 8; Length 800;
Best Local Similarity	87.6%; Pred. No. 6,4e-56;
Matches 700; Conservative	0; Mismatches 90; Indels 9; Gaps 2;
0Y	512 cggcgccgagcgacactgtccgcccgaatccgtgtaggatacccaagcagatgacgactgagc 571
Db	1 CGCCGTCGCGCGACCCGCGACGCGACGCGCGCGCGCGGAGGCCCAAGCGGATGAGACTGGCG 60
0Y	572 ggaagcagactcgtcgatcatctctctctctcgtgctctcgtgtggg--gcgcacaga 628
Db	61 GCGGACGACGTCGTCCTCTTCCTGTCGTATCATGATGATGCGGCGATGCACATAGGA 120
0Y	629 gctcgtggtggaagctgcccgcgcgg-----tcggcgccgagcgaagcagcgcgcgc 682
Db	122 GCTAGTGTGTCGAACACTCGCGCGCGCGGCGGACGACCAAGCATCCGCGCGGCGCAACGGGCCG 180
0Y	683 cgttcgcgtgctgctgagtcgacacgacgagggccggagattcgctgtgtgacgcgcctgc 742

Db	181	CGGTCCGGTGTGTTGTGTGAACAGCAGAGAGGCCGGATCCGGCTCGGCACGGCGTGTC	240
Oy	743	tgtggttcgcgggaagccgtctcaagaagagaacctctcccgcgcggaaggcgtcgttgaagc	802
Db	241	TGGGGTGC GGAGGCCGTGCACAGAGAACAATTCTCTCGGGGGAGGGCGCTGTGTCAAGC	300
Oy	803	agataaccttgcttgcgcgcgtgccaaaggcgcgcgatgcgcgaagtctgcgcctaactccg	862
Db	301	AGATCCCCAATGCTGGCTCTCTGTCAGGGGGGTGCCATGCGCAAGGTCGGCCGCTACTTCCG	360
Oy	863	ggcgaaggccctcgcgcgcgcgcttccgcgtctccgcgcgcgaagcgaagctccccctccg	922
Db	361	GCGAGGGCGCTTGCCCGCGCGGTGTATCGCTTCCGCCCGCACCGGAACAECTCCCTCTCCG	420
Oy	923	accgcgccttcgcgcgaactctctccaacgcgcgcgaattctcaagttcccttaactcaagt	982
Db	421	ACGGCGGCTTTCGCGAGACTCTTGTGACGGCGGACACTTTCAGCAAGTCTCTGCCCTCATCGAAGT	480
Oy	983	tcgcgcacttcaaccgcaccaacgaagcatctctgaaggcgttcgcgcgcgtcgcgcgcgtgc	1042
Db	481	TGCGCCCACTTTCACCGCGCAACCAAGGCCATCTCCAGGCGCTTCCCGGCGTCCCGCCGCTCC	540
Oy	1043	acgtctcgtcgaattcggcataaagcaaggatgtgaatgagccgcgaattctccaaggccctcg	1102
Db	541	ACGTGCTGTAATTTCGTGATCAACACAGGGGATGACATGGCGGCTCTTCTTCACAGGCCCTCG	600
Oy	1103	ccctccgcctccgcgcgcgcctccctctcgtcttccgcctaacgcgcgcgtcgcgcgcgcgcgcg	1162
Db	601	CCTCTCCGCCCTCGGCGGCCCGCCCGCTCGTCCGCGCTCACCGGGCTCGGCGCGGACGCCG	660
Oy	1163	acgaagaccgaacgcctctgcacaagatgtagtgaagctcgcgcacagltcgcgcacacattcc	1222
Db	661	ACGAGACCGAAGCCTTGCACAGAGTAGGGGCTGAAACTTGCCACAGTTCCGCGCACACATACC	720
Oy	1223	ggcttcgaattccagtaaccgcgcgcctcgtcgcgcgcacagctcgcgcgaactctggaagccgttca	1282
Db	721	GCGTGGACTTTCACATACCGTGGCGCTGTCGCGGCCACAGCTCGCCGACNCGTGGAGGCCGTCCA	780
Oy	1283	tgtctgcagcgcgggaaggcga	1301
Db	781	TGCTGCAACCGGAGGGCGCA	799
RESULT	15		
LOCUS	AF377645	800 bp DNA	PLN 01-AUG-2001
DEFINITION	Zea mays subsp. mays isolate B73 gibberellin response modulator gene, partial cds.		
ACCESSION	AF377645		
VERSION	AF377645.1	GI:14318162	
KEYWORDS	maize;		
ORGANISM	Zea mays subsp. mays		
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.		
REFERENCE	1 (bases 1 to 800)		
TITLE	Tenailion,M.I., Sawkins,M.C., Long,A.D., Gaut,R.L., Doebley,J.F. and Gaut,B.S.		
JOURNAL	Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.)		
PUBMED	Proc. Natl. Acad. Sci. U.S.A. 98 (16), 9161-9166 (2001).		
REFERENCE	2 (bases 1 to 800)		
AUTHORS	Tenailion,M.I., Sawkins,M.C., Gaut,R.L., Long,A.D., Doebley,J.F. and Gaut,B.S.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-MAY-2001) Ecology and Evolutionary Biology,		
FEATURES	University of California, Irvine, 321 steinhans, Irvine, CA 92697-2525, USA		
SOURCE	Location/Qualifiers		
	1..800		









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2002, 00:58:16 ; Search time 197.56 Seconds  
(without alignments)  
9221.584 Million cell updates/sec

Title: US-09-485-529-14

Perfect score: 2125

Sequence: 1 atagagagggagtagctc.....tggatgagcagcaactccg 2125

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq.1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	2125	20	AA336279
2	1956.2	92.1	2709	20	AA336275
3	1502.4	70.7	1768	20	AA336278
4	1379.8	64.9	1746	20	AA336285
5	1317	62.0	2255	20	AA336280
6	420.4	19.8	770	20	AA336277
7	382	18.0	1602	21	AA345745
8	381.8	18.0	1964	18	AAT91937
9	380.2	17.9	1951	22	AAD06646
10	380.2	17.9	1951	22	AAD05791
11	380	17.9	22	AAF25480	

12	378.4	17.8	1779	22	AAF25481
13	367	17.3	416	20	AA336283
14	359	16.9	436	20	AA336269
15	351.8	16.6	1643	18	AAT91938
16	347	16.3	1642	18	AAT91941
17	343	16.1	1642	18	AAT91939
18	329.4	15.5	1636	18	AA336240
19	329.2	15.5	725	20	AA336276
20	326.6	15.4	1764	22	AAD06661
21	326.6	15.4	1764	22	AAD05776
22	322	15.2	377	20	AA336268
23	316.6	14.9	369	20	AA336258
24	312.2	14.7	453	20	AA336263
25	308	14.5	332	20	AA336255
26	297	14.0	324	20	AA336273
27	277.6	13.1	357	20	AA336259
28	267	12.6	425	20	AA336270
29	264.2	12.4	377	20	AA336256
30	249.2	11.7	309	20	AA336261
31	242	11.4	285	20	AA336271
32	238.4	11.2	511	20	AA336260
33	180	8.5	497	21	AA336267
34	178.4	8.4	230	20	AA336272
35	174.8	8.2	2151	18	AAT95767
36	174.8	8.2	2151	21	AA336298
37	171.8	8.1	3510	21	AA336299
38	171.6	8.1	359	20	AA336262
39	169.2	8.0	259	20	AA336272
40	165.8	7.6	302	20	AA336281
41	161	7.6	211	20	AA336257
42	155	7.3	969	21	AA336327
43	153.8	7.2	341	21	AA336316
44	153.8	7.2	341	21	AA336358
45	148	7.0	371	20	AA336282

#### ALIGNMENTS

RESULT 1  
ID AAX36279 standard; DNA; 2125 BP.  
XX AAX36279;  
XX  
AC AAX36279;  
XX  
DT 16-JUL-1999 (first entry)  
XX  
DE Wheat Rht clone 5a1 genomic sequence.  
XX  
KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
KW paclobutrazol; ss.  
XX  
OS Triticum aestivum.  
XX  
PN WO9909174-A1.  
XX  
PD 25-FEB-1999.  
XX  
PF 07-AUG-1998; 98WO-GB02383.  
XX  
PR 13-AUG-1997; 97GB-0017192.  
XX  
PA (PLAN-) PLANT BIOSCIENCE LTD.  
XX  
PI Harberd NP, Peng J, Richards DE;  
XX WPI: 1999-181040/15.  
XX P-PSDB: AAT02540.  
XX  
PT New Triticum aestivum polynucleotides - encode a polypeptide which provides inhibition of the growth of plants, which inhibition is antagonised by gibberellin, used to confer a dwarf phenotype

Nucleotide sequenc  
Partial sequence o  
DNA sequence obtai  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Partial cDNA seque  
A. thaliana transc  
Arabidopsis thalia  
DNA sequence obtai  
DNA sequence obtai  
DNA sequence obtai  
DNA sequence obtai  
DNA sequence obtai  
DNA sequence obtai  
DNA sequence obtai  
DNA sequence obtai  
DNA sequence obtai  
DNA sequence obtai  
Pinus radiata tran  
DNA sequence obtai  
Maize ZCARECROW ZC  
Maize ZCR gene par  
Maize Scarecrow nu  
DNA sequence obtai  
DNA sequence obtai  
Partial sequence o  
DNA sequence obtai  
Maize GBBHT74 par  
Eucalyptus grandis  
Eucalyptus grandis  
Partial sequence o

XX Disclosure: Fig 8a; 88pp; English.  
PS The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in Triticum aestivum, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence represents  
CC the wheat Rht clone 5a1 genomic sequence.  
XX  
Sequence 2125 BP: 325 A; 768 C; 723 G; 309 T; 0 other;

Query Match 100.0%; Score 2125; DB 20; Length 2125;  
Best Local Similarity 100.0%; Pred. No. 1,1e-282;  
Matches 2125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagagagcgaggtagctcgcgagatcatalgaagcggaagtaccagagcgcgagagagc 60  
DB 1 atagagagcgaggtagctcgcgagatcatalgaagcggaagtaccagagcgcgagagagc 60  
QY 61 ggcgagcgagctgagcgagatggtctgcgcagagagaatgagtctcgcgagcgcg 120  
DB 61 ggcgagcgagctgagcgagatggtctgcgcagagagaatgagtctcgcgagcgcg 120  
QY 121 ggggagggggaaggtgagacagatgctgagcgagcgctcgaggtataaaggtgagcgctcc 180  
DB 121 ggggagggggaaggtgagacagatgctgagcgagcgctcgaggtataaaggtgagcgctcc 180  
QY 181 gacacgtgagagatgagcgagagagcttgagacagctcgagatgagcattgaggtgagcgcg 240  
DB 181 gacacgtgagagatgagcgagagagcttgagacagctcgagatgagcattgaggtgagcgcg 240  
QY 241 gtaggagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 300  
DB 241 gtaggagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 300  
QY 301 gactcaaacacccacagcactgctgctgaggtcgagagacatgctctcgagagctcaacgcg 360  
DB 301 gactcaaacacccacagcactgctgctgaggtcgagagacatgctctcgagagctcaacgcg 360  
QY 361 ccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 420  
DB 361 ccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 420  
QY 421 ggcagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 480  
DB 421 ggcagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 480  
QY 481 ctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 540  
DB 481 ctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 540  
QY 541 gtagcgagatcccaagcgagatgagcactgagcgagagcagacactctgctcatcctctcc 600  
DB 541 gtagcgagatcccaagcgagatgagcactgagcgagagcagacactctgctcatcctctcc 600  
QY 601 tctgctctctctctgtaggggagcgagagctctgtagtagagcgagcgagcgagcgagcgagcg 660  
DB 601 tctgctctctctctgtaggggagcgagagctctgtagtagagcgagcgagcgagcgagcgagcg 660  
QY 661 ggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 720  
DB 661 ggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 720

QY 721 atcgagctgtagcagcgagctgctgagctgagcgagcgagcgagcgagcgagcgagcgagcgagcg 780  
DB 721 atcgagctgtagcagcgagctgctgagctgagcgagcgagcgagcgagcgagcgagcgagcgagcg 780  
QY 781 gccgagagcgagctgtagaagagataccctgagcgagcgagcgagcgagcgagcgagcgagcgag 840  
DB 781 gccgagagcgagctgtagaagagataccctgagcgagcgagcgagcgagcgagcgagcgagcgag 840  
QY 841 cgaagagctcgagcgagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 900  
DB 841 cgaagagctcgagcgagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 900  
QY 901 cagcgagagcgagctcctctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 960  
DB 901 cagcgagagcgagctcctctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 960  
QY 961 gagctcgcctcctacctaagtctgagcgagcttcagcgagcgagcgagcgagcgagcgagcgagcg 1020  
DB 961 gagctcgcctcctacctaagtctgagcgagcttcagcgagcgagcgagcgagcgagcgagcgagcg 1020  
QY 1021 ttcgagcgagctgagcgagcgagctgagcgagcttcagcgagcgagcgagcgagcgagcgagcgag 1080  
DB 1021 ttcgagcgagctgagcgagcgagctgagcgagcttcagcgagcgagcgagcgagcgagcgagcgag 1080  
QY 1081 ccgagacattctcagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1140  
DB 1081 ccgagacattctcagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1140  
QY 1141 ggcgtcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1200  
DB 1141 ggcgtcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1200  
QY 1201 gccagagctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1260  
DB 1201 gccagagctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1260  
QY 1261 ctcgagcgagctgagcgagctcagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1320  
DB 1261 ctcgagcgagctgagcgagctcagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1320  
QY 1321 ccgagagctatcgagcgagctcagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1380  
DB 1321 ccgagagctatcgagcgagctcagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1380  
QY 1381 gccctgagagagctcctgagcgagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1440  
DB 1381 gccctgagagagctcctgagcgagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1440  
QY 1441 gagcagagagcgagctcagcgagctcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1500  
DB 1441 gagcagagagcgagctcagcgagctcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1500  
QY 1501 tactactcaacagcttgcagctcctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1560  
DB 1501 tactactcaacagcttgcagctcctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1560  
QY 1561 gctcatatcgagcgagctcctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1620  
DB 1561 gctcatatcgagcgagctcctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1620  
QY 1621 gtagtaacttcgagcgagctcagcgagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1680  
DB 1621 gtagtaacttcgagcgagctcagcgagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1680  
QY 1681 cgcagagagagcgagctgagcgagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1740  
DB 1681 cgcagagagagcgagctgagcgagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1740  
QY 1741 cagctgagctcagctcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1800  
DB 1741 cagctgagctcagctcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1800  
QY 1801 gacggctacaaggtgagagagagagcgagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1860

	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2
--	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	---

## RESULT 2

ID AAX36275 standard; DNA; 2709 BP.

AC AAX36275;

DT 16-JUL-1999 (first entry)

Composite DNA sequence of wheat Rht gene.

KW Rht gene; homologue; *Triticum aestivum*; wheat; growth inhibition;  
 KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
 KW paclobutrazol; ss.

**Triticum aestivum.**

PN WO9909174-A1

PD 25-FEB-1999.

PF 07-AUG-1998; 98WO-GB02383.

PR 13-AUG-1997; 97GB-0017192.

PA (PLAN-) PLANT BIOSCIENCE LTD

PI Harberd NP, Peng J, Richards DE;

DR WPI; 1999-181040/15.

PT New Triticum Aestivum polynucleotides - encode a polypeptide which

PT antagonised by gibberellin, used to confer a dwarf phenotype

PS Disclosure; Fig 3a; 88pp; English.

CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in *Triticum aestivum*, inhibit  
CC growth of the plant. This growth inhibition is antagonized by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence represents

CC the composite DNA sequence of wheat Rht gene.  
XX  
SQ Sequence 2709 BP; 473 A; 900 C; 823 G; 482 T; 31 other,

Query Match	92.1%;	Score 1956.2;	DB 20;	Length 2709;
Best Local Similarity	96.8%;	Pred. No. 1.3e-259;		
Matches 2061; Conservative	0;	Mismatches 53;	Indels 15;	Gaps 77;

[illegible]

[illegible]

```

Db      2164  cgcctgattgcagcgaccgcgctactgcatcgagctcaggttgctccgctcgttcgc 2223
QY      2097  gtgaagaggttgatgacgacgaactccg 2125
Db      2224  gtgaagaggttgatgacgacgaactccg 2252
RESULT  3
ID      AAX36278
XX      AAX36278 standard; cDNA; 1768 BP.
AC      AAX36278;
XX      16-JUL-1999 (first entry)
DT      16-JUL-1999 (first entry)
XX      Wheat Rht clone Cl5-1 cDNA sequence.
DE      Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
KW      antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KM      paclobutrazol; ss.
XX      Triticum aestivum.
OS      Triticum aestivum.
XX      WO9909174-A1.
PM      25-FEB-1999.
PD      07-AUG-1998; 98WO-GB02383.
PF      13-AUG-1997; 97GB-0017192.
XX      (PLAN-) PLANT BIOSCIENCE LTD.
PA      Harberd NP, Peng J, Richards DE;
XX      WPI: 1999-181040/15.
XX      P-PSDB; AAY02339.
DR      New Triticum Aestivum polynucleotides - encode a polypeptide which
PT      provides inhibition of the growth of plants, which inhibition is
PT      antagonised by gibberellin, used to confer a dwarf phenotype
XX      Disclosure: Fig 7a; 86pp; English.
XX      The specification describes polypeptides encoded by the Rht gene (and
XX      its homologues) that, when expressed in Triticum Aestivum, inhibit
XX      growth of the plant. This growth inhibition is antagonised by
XX      gibberellin. The products can be used to provide Rht expression in
XX      plants, conferring a dwarf phenotype on a plant which is correctable
XX      by treatment with gibberellin. In addition, the products can be
XX      used to produce Rht mutant plants which are dwarfed compared with
XX      wild-type, the dwarfing being gibberellin-insensitive. Taller plants
XX      may be made by knocking out Rht or the relevant homologous gene in
XX      the plant of interest. Plants may be made which are resistant to
XX      compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
XX      e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
XX      dwarf but let crop plants grow tall. The present sequence represents
XX      the wheat Rht clone Cl5-1 cDNA sequence.
SQ      Sequence 1768 BP; 368 A; 595 C; 527 G; 278 T; 0 other;
Query Match      70.7%; Score 1502.4; DB 20; Length 1768;
Best Local Similarity 99.9%; Pred. NO. 1.3e-197;
Matches 1503; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      622  gccagagactctgtgtgtgagagctgcccgccgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 681
Db      1  gccagagactctgtgtgtgagagctgcccgccgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 60
QY      682  gccgtccgctcgtctgtgtgcacacgacgacgacgacgacgacgacgacgacgacgacgacgac 741

```

[illegible]

QY	1872	aaggaaggctctgcctgaagctcggggtggaacacgcgcgcgcctgtatcgccacctggcattg	1881
Db	1201	aaggaaggctctccctgacgctgtgggtggcaacacgcgcgcctgtatcgccacctggcattg	1260
QY	1882	cgccctgcccggccgtatctcgcgagttttgaacgcctgtcaagtaacacacgttgacatg	1941
Db	1261	cgcttgccgggcccgtatctcgcgagttttgaacgcctgtcaagtaacacacgttgacatg	1320
QY	1942	gaggacaacacagcccccgcgcgcgcgcgcctctccgcgcgaacgcacgcacgcgcac	2001
Db	1321	gaggacaacacacagcccccgcgcgcgcgcctctccgcgcgaacgcacgcacgcgcac	1380
QY	2002	ttgaagaagaagaagcctaatagtcatgttcagtgagcgtcaattgcagcgcgcctacg	2061
Db	1381	ttgaagaagaagaagcctaatagtcatgttcagtgagcgtcaattgcagcgcgcctacg	1440
QY	2062	atcagtcggagtaagggtgtgttcctccgtcttgacgctgaagaaggttgatggaacgaac	2121
Db	1441	atcagtcggagtaagggtgtgttcctccgtcttgacgctgaagaaggttgatggaacgaac	1500
QY	2122	tcgcg 2125	
Db	1501	tcgcg 1504	
RESULT 4			
ID	AA36285	standard; cDNA; 1746 BP.	
XX	AA36285		
AC	AA36285		
XX			
DT	16-JUL-1999	(first entry)	
XX			
DE		Consensus cDNA sequence of wheat Rht clone C15-1.	
XX			
KW		Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;	
KW		antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;	
KW		paclobutrazol; ss.	
XX			
OS		Triticum aestivum.	
XX			
PN	W09909174-A1.		
XX			
PD	25-FEB-1999.		
XX			
PF	07-AUG-1998;	98WO-GB02383.	
XX			
PR	13-AUG-1997;	97GB-0017192.	
XX			
PA	(PLAN-)	PLANT BIOSCIENCE LTD.	
XX			
PI	Harberd NP, Peng J, Richards DE;		
XX			
XR	WPI; 1999-181040/15.		
XX			
CC	The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in Triticum Aestivum, inhibit growth of the plant. This growth inhibition is antagonised by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol.		





D	b	156	cgcgcgcggcgcccttcgccgagcgggtacagggtggaagaagaacggcgcc
O	y	1835	tgcagcttgagggtggcacacgcgcgcgtgatcgcacctcgagcatgctgcgcg
D	b	2016	tgaccctcgagggtggcatacgcgcgcctcatcgccaactcgcgcgtgcgcgcg
<hr/>			
RESULT 6			
A	A	X	36277 standard; DNM: 770 BP.
A	A	X	36277;
X	X	X	AA36277;
D	T	T	16-JUL-1999 (first entry)
D	E	Rice EST D39460 sequence, homologous to wheat Rht gene.	
X	X	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;	
K	M	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;	
X	X	pachlobutrazol; rice; expressed sequence tag; EST, ss.	
X	S	Oryza sativa.	
X	N	WO9909174-AI.	
P	D	25-FEB-1999.	
P	F	07-AUG-1998; 98MO-GB02383.	
X	R	13-AUG-1997; 97GB-0017192.	
P	A	(PLAN-) PLANT BIOSCIENCE LTD.	
X	X	Harberd NP, Peng J, Richards DE;	
X	D	WPI; 1999-181040/15.	
D	R	P-PDSB; AAY02538.	
P	T	New Triticum aestivum polynucleotides - encode a polypeptide which	
P	T	provides inhibition of the growth of plants, which inhibition is	
X	X	antagonised by gibberellin, used to confer a dwarf phenotype	
X	P	Claim 13; Fig 6a; 88pp; English.	
C	C	The specification describes polypeptides encoded by the Rht gene (and	
C	C	its homologues) that, when expressed in Triticum aestivum, inhibit	
C	C	growth of the plant. This growth inhibition is antagonised by	
C	C	gibberellin. The products can be used to provide Rht expression in	
C	C	plants, conferring a dwarf phenotype on a plant which is correctable	
C	C	by treatment with gibberellin. In addition, the products can be	
C	C	used to produce Rht mutant plants which are dwarfed compared with	
C	C	wild-type, the dwarfing being gibberellin-insensitive. Taller plants	
C	C	may be made by knocking out Rht or the relevant homologous gene in	
C	C	the plant of interest. Plants may be made which are resistant to	
C	C	compounds which inhibit gibberellin biosynthesis, such as pachlobutrazo	
C	C	e.g., to allow use of a gibberellin biosynthesis inhibitor to keep weed	
C	C	dwarf but let crop plants grow tall. The present sequence represents	
C	C	rice expressed sequence tag (EST) AAD39460, which is homologous to the	
C	C	wheat Rht gene.	
X	X	Sequence 770 BP; 104 A; 252 C; 294 G; 130 T; 0 other;	

Query Match	19.8%;	Score 420.4;	DB 20;	Length 770;
Best Local Similarity	77.5%;	Pred. No. 9.1e-50;		
Matches 595;	Conservative	0;	Mismatches 146;	Indels 27; Gaps 6;

[illegible][illegible]

RESULT	7
AAC45745	
ID	AAC45745 standard; DNA; 1602 BP.

AC AAC45745;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 47609.

KM Hybridisation assay; genetic mapping; gene expression control;  
KM protein identification; signal transduction pathway;  
KM metabolic pathway; promoter; termination sequence; ss.

OS *Arabidopsis thaliana*.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 09-MAR-1999; 99US-0123548.



PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142877.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151338.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.

```

PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

```

Query Match 18.0%; Score 382; DB 21; Length 1602;

Best Local Similarity 60.0%; Pred. No. 1.4e-44; Matches 736; Conservative 0; Mismatches 415; Indels 75; Gaps 3;

```

QY 665 cggcacaacgagccgcgcgtcgctgctgctgcacacgagagccgagatc 724
DB 443 caggcagcgatgataactcgcgatgtctcgttgcacgcgagagcggtg 502
QY 725 ggcctgtagcgcgcgtcgctgctgagcgagcgctgacagagaaacctc 784
DB 503 gctcgtcacaacgcgtcttgctgctgcgcgagcgtgtaagaagaagaa 562
QY 785 cggagggcgctgtagaagacataaccttgctgacgcgcgtcccaaggcg 844
DB 563 cggaaagctctgtgtaagaacaaatcgaattcttaagccgttctcaaat 622
QY 845 aggtcgcgcgtactctgagagagccctgcgcgcgcgtcttcgcgcttc 904
DB 623 aagtcgctactactctgcgcgcgcgtctgcgcgcgcgtcttcgcgct 682
QY 905 cggacaagcctcctctcgcgcgcgcgtctgcgcgcgcgtcttcgcgc 964
DB 683 ag-----agtcacaacgcacactcctcgcgatacttcaagatgca 736
QY 965 cctgcaccctaccatgaattcgcgcacattcaaccgcaaccagagccalc 1024
DB 737 ctgtccctatctcaagtctcgaacttcaacgagcaatcaagcgattcttc 796
QY 1025 ccgagctgcgcgcggtgcaagctgcgcgcgtctgcgcacaaagagggat 1084
DB 797 aagggaagaagaagattctcatgtatcttctatgtagtcaaggtcttcaa 856
QY 1085 caactctcagaagcctcgcgcgcgtctccgcgcgcgcgtcttcgcgc 1144
DB 857 cgcctatgcaaggctctcgcgcgcgcgtccgcgcgcgtcttcgcgcgc 916
QY 1145 tcggccccccgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1204
DB 917 ttggtcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 976
QY 1205 agtgcgcgcacacatccgcgtcgcgcgcgcgcgcgcgcgcgcgcgc 1264

```

```

DB 977 atttagctgagcgatcaactcagcttgatgtagtacaagagatttgccta 1036
QY 1265 cggacctggaagccgttcttcgctgcgcgcgcgcgcgcgcgcgcgcgcgc 1324
DB 1037 cggactctgtagctcgaagcttgagc-----ttgacaagtgaaatg 1081
QY 1325 aggtatccgcctcaactcaactcagcttcgagatgacacgcgcgtctgcgc 1384
DB 1082 aatcgttgctggttaactctgttcttcgcgcgcgcgcgcgcgcgcgcgcgc 1141
QY 1385 tggagaaggtctctgagcaacgctgcgcgcgcgcgcgcgcgcgcgcgcgc 1444
DB 1142 tcgataaagttctctgtgtgtgtaacagatcaaacgcgcgcgcgcgcgcgc 1201
QY 1445 aggaagcgcaatcacacactccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1504
DB 1202 aggaatcgacaataatgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1261
QY 1505 actccacacatgttcgattccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1564
DB 1262 actcgacgtgtgttgaacctcgttggaag-----1289
QY 1565 catcggggctgctgctcctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1624
DB 1290 -----tgtaaccgagtgatcaagaacagatcatgtcgcgcgcgcgcgc 1327
QY 1625 acctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1684
DB 1328 acttggttaaacaagatctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1387
QY 1685 acgagacgctgagcgagtgcgagacgcgcgcgcgcgcgcgcgcgcgcgcgc 1744
DB 1388 atgaacacgttgtagtaagtgaggaacgcgcgcgcgcgcgcgcgcgcgcgcgc 1447
QY 1745 tgggtctcaatgctcaacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1804
DB 1448 ttggtctgaatcgcttgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1507
QY 1805 gctacaaggttgaggaaggaagagcgtcgcgcgcgcgcgcgcgcgcgcgcgc 1864
DB 1508 gttacgcgggttgaggaagagtgagcgcgcgcgcgcgcgcgcgcgcgcgcgc 1567
QY 1865 tcgcacactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1890
DB 1568 tagccacactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1593

RESULT 8
AAT91937
ID AAT91937 standard; DNA; 1964 BP.
XX
AC AAT91937;
XX
DT 19-MAR-1998 (first entry)
XX
DE Arabidopsis thaliana gibberellin insensitivity gene gal.
XX
KW Gibberellin insensitivity; gal; plant growth inhibition;
KW dwarf phenotype; lodging resistance; increased yield;
KW flowering regulation; bolting inhibition; spinach; lettuce;
KW antibody; identification; probe; primer; antisense; sense;
KW expression regulation; co-suppression; rice;
KW Bakane disease resistance; ss.
XX
OS Arabidopsis thaliana.
XX
PN W09729123-A2.
XX
PD 14-AUG-1997.
XX
PF 12-FEB-1997; 97WO-GB00390.
XX

```







DB 1517 ttacttggtaaacagatcgaacgttctgtgtgtatgtagctgacccgagttgagc 1576  
QY 1682 gccacgagcgctggcgacgttgcggaacggcttggcaacgggttcgagccgtcc 1741  
DB 1577 gtaataaacgtttgagtcagtggaacggcttcggtctcgtgttctgcgcgcgac 1636  
QY 1742 acctggcctccatcgaacgaacgagcagcagcgtctgtgcgtcttcgcgcgcgag 1801  
DB 1637 atattgttcgaatgcgtttaagcaagcgagtatgtcttgcgtcttcgtaacagcggtg 1696  
QY 1802 acggtcacaagtgtagaggaaggaagcgtcgtcgaacgttggtggtgcaacgcgcgc 1861  
DB 1697 aggtgtatcgtgtgtagagagtagcagcgtgtctcatgttggttggtagcaacgcgcgc 1756  
QY 1862 tgatcgccacgtcgatggcgctggcc 1890  
DB 1757 tcatagcacacccgcgtcttggaaactctcc 1785

## RESULT 11

AAF25480

ID AAF25480 standard: DNA; 1779 BP.

AC AAF25480;

DT 15-MAY-2001 (first entry)

DE Nucleotide sequence of a wildtype GRAS protein.

KW GRAS protein; BZH gene; transcription factor; gibberellin; morphogenesis;  
KW plant development; dwarf plant; crucifer; ss.

OS Brassica napus.

FH Key Location/Qualifiers

FT CDS 60..1778

FT /tag= "a"

FT /product= "GRAS"

PN MO200109356-A1.

PD 08-FEB-2001.

PF 02-AUG-2000; 2000MO-FR02216.

PR 02-AUG-1999; 99FR-0010023.

PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.

PI Renard M, Delourme R, Barret P, Brunel D, Froger N, Tanguy X;

DR MPI; 2001-182964/18.

DR P-PSDB; AAB31883.

PT New mutant nucleic acid encoding modified GRAS family protein, used to  
PT produce dwarf transgenic plants

PS Example 1; Page 13-15; 28pp; French.

CC The present sequence encodes a wild type plant protein of the GRAS  
CC family. The specification describes a mutant allele of the BZH gene,  
CC which contains a G1695A mutation resulting in the mutation E546K in the  
CC protein. GRAS proteins are transcription factors implicated in  
CC regulation of the response to gibberellins and thus in control of  
CC morphogenesis and plant development. The mutant GRAS protein is  
CC used to produce dwarf plants, specifically crucifers. Dwarf plants may  
CC be sown earlier (increasing nitrate accumulation without risking  
CC excessive stem growth during winter), and have better resistance to  
CC cold and lodging. They are also easier to harvest and allow for better  
CC monitoring of the crop.

SQ Sequence 1779 BP; 418 A; 430 C; 476 G; 455 T; 0 other;

Query Match 17.9%; Score 380; DB 22; Length 1779;  
Best Local Similarity 60.1%; Pred. No. 2,7e-44;  
Matches 731; Conservative 0; Mismatches 410; Indels 75; Gaps 3;

QY 678 gcccgctgcgcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 737  
DB 629 gcaactcgtcttcaactcgttcgttcgacagcagagcgaggttcgttcgacgc 688  
QY 738 gcttcgttcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 797  
DB 689 gcttcgttcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 748  
QY 798 gaagcagatcccttcgttcgttcgacagcagagcgaggttcgttcgacgc 857  
DB 749 taagcagatcccttcgttcgttcgacagcagagcgaggttcgttcgacgc 808  
QY 858 ctccgagagcgcccttcgttcgttcgacagcagagcgaggttcgttcgacgc 917  
DB 809 ctccgagagcgcccttcgttcgttcgacagcagagcgaggttcgttcgacgc 862  
QY 918 ctccgagagcgcccttcgttcgttcgacagcagagcgaggttcgttcgacgc 977  
DB 863 gacgacatcccttcgttcgttcgacagcagagcgaggttcgttcgacgc 922  
QY 978 caagtcgcgacatcccttcgttcgttcgacagcagagcgaggttcgttcgacgc 1037  
DB 923 caagtcgcgacatcccttcgttcgttcgacagcagagcgaggttcgttcgacgc 982  
QY 1038 cgttcgagtcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1097  
DB 983 agtcgagtcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1042  
QY 1098 ccttcgagtcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1157  
DB 1043 ccttcgagtcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1102  
QY 1158 gccgagtcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1217  
DB 1103 gccgagtcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1162  
QY 1218 catcgttcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1277  
DB 1163 gattcagtcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1222  
QY 1278 gttcagtcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1337  
DB 1223 cttcagtcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1267  
QY 1338 caactcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1397  
DB 1268 taactcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1327  
QY 1398 gggcagtcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1457  
DB 1328 cggcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1387  
QY 1458 caactcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1517  
DB 1388 taactcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1447  
QY 1518 cgttcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1577  
DB 1448 tttcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1462  
QY 1578 tttcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1637  
DB 1463 tttcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1513  
QY 1638 gattcagtcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1697  
DB 1514 gattcagtcgttcgttcgttcgacagcagagcgaggttcgttcgacgc 1573

QY	1698	ccgtgctggaacccgctggtggaacgcgcgggttcgagacgtccacactggtctccatgac	1757
Db	1574	tcatagtctgaacccggttcggttcgctgcgggttttcgcgcgcgcacatctcgggtctaacgc	1633
QY	1758	ctcaaacgaagcgagcacgcctgtctgtgcgtcttcgcgcgcgcgcgcgcgtctaacaggtgga	1817
Db	1634	gtttaagcaaacgagcaacgcttttgcgtcttcttaatggaaggcgaaggtacggtgtgga	1693
QY	1818	ggaagaagaagctgtcctgacgtctggtggtgagacaacgcgcgcgcgtgacgcgaactcgac	1877
Db	1694	ggagaaataatgggtgtgttgatgtgtgattggtgcgaacactgcaccgcgtcataaccactcgc	1753
QY	1878	atggcgccctggtgcccggg	1893
Db	1754	ttggaagctctcgcg	1769
RESULT 12			
AAAF25481	ID	AAAF25481 standard; DNA; 1779 BP.	
AAAF25481.	AC		
AAAF25481.	AF		
15-MAY-2001	DF	(first entry)	
Nucleotide sequence of a mutant GRAS protein.	DE		
GRAS protein; B2H gene; transcription factor; gibberellin; morphogenesis;	KW		
plant development; dwarf plant; crucifer; ss.	KM		
Brassica napus.	OS		
Key	EH	Location/Qualifiers	
CDS	FT	60..1778	
/*tag=	FT	a	
/product=	FT	"GRAS"	
MO200109356-A1.	PN		
08-FEB-2001.	PD		
02-AUG-2000; 2000MO-FR02216.	PE		
02-AUG-1999; 99FR-0010023.	PR		
(INRG ) INRA INST NAT RECH AGRONOMIQUE.	PA		
Renard M, Delourme R, Barret P, Brunel D, Froger N, Tanguy X;	PI		
MP1; 2001-182964/18.	DR		
P-PSDB; AAB31884.	DR		
New mutant nucleic acid encoding modified GRAS family protein, used to	PT		
produce dwarf transgenic plants	PT		
Example 1; Page 18-20; 28pp; French.	PS		
The present sequence encodes a mutant plant protein of the GRAS	CC		
family. The mutant allele of the B2H gene contains a G1695A mutation	CC		
resulting in the mutation E546K in the protein. GRAS proteins are	CC		
transcription factors implicated in regulation of the response to	CC		
gibberellins and thus in control of morphogenesis and plant development.	CC		
The mutant GRAS protein is used to produce dwarf plants, specifically	CC		
crucifers. Dwarf plants may be sown earlier (increasing nitrate	CC		
accumulation without risking excessive stem growth during winter), and	CC		
have better resistance to cold and lodging. They are also easier to	CC		
harvest and allow for better monitoring of the crop.	CC		
Sequence 1779 BP; 419 A; 430 C; 475 G; 455 T; 0 other.	SO		

Query Match

17.8%; Score 378.4; DB 22; Length 1779;

Best Local Similarity

60.0%; Pred No. 4.4e-44;

[illegible]

OY	1758	ctcaacagcaggcgagacaagcgtgcgaccttccttccggcggcgcgaacgctaagaagtgcga	1817		
Xx					
Db	1634	glcttaagaaagcagatacagactcttttgctcttctttaatgatgagcgaaaagtattcatgttgtga	1693		
Xx					
OY	1818	ggaagaagaaagcctbcgccgacgctctgggtgcgacacagcgcgcctgatatgcaccactgcgc	1877		
Xx					
Db	1694	gaagataaatgtggtgttgtagtcttgtagtctgcacactcagcacctcataaacacactccgc	1753		
Xx					
OY	1878	atggcgccttgagccggg	1893		
Xx					
Db	1754	ttygaagctctcgcg	1769		
Xx					
<b>RESULT 13</b>					
ID	AAX36283				
Xx	AAX36283 standard; DNA; 416 BP.				
AC					
Xx	AAX36283;				
DT					
Xx	16-JUL-1999 (first entry)				
DE					
Pf	Partial sequence of the wheat rht-10 allele.				
Rw					
KW	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;				
Xx	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;				
Kv	paclobutrazol; ss.				
Xx					
OS	Triticum aestivum.				
Xx					
PN	M09909174-AI.				
Xx					
PD	25-FEB-1999.				
Xx					
PF	07-AUG-1998; 98MO-G802383.				
Xx					
PR	13-AUG-1997; 97GB-0017192.				
Xx					
PA	(PLAN-) PLANT BIOSCIENCE LTD.				
Xx					
PI	Harberd NP, Peng J, Richards DE;				
Xx					
DR	WPI: 1999-181040/15.				
Xx					
PT	P-PSTDB; MAY02544.				
Xx					
XX	New Triticum Aestivum polynucleotides - encode a polypeptide which provides inhibition of the growth of plants, which inhibition is antagonised by gibberellin, used to confer a dwarf phenotype				
XX	Disclosure; Fig 12a; 88bp; English.				
CC	The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in Triticum Aestivum, inhibit growth of the plant. This growth inhibition is antagonised by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds dwarf but let crop plants grow tall. The present sequence represents the partial sequence of the wheat rht-10 allele.				
Sequence	416 BP; 59 A; 168 C; 129 G; 60 T; 0 other;				

Query Match	17.3%;	Score 367;	DB 20;	Length 416;
Best Local Similarity	91.4%;	Pred. No. 1.9e-42;		
Matches 416;	Conservative 0;	Mismatches 0;	Indels 39;	Gaps 1;

OY	151	g c g c g c t c g g g t a c a a g g t g c g c g c t c c c g a c a t a t g c g a g a c t g n g c g a g a a g c t g n g a g	210
Db	1	g c g c g c g c t c g g g t a c a a g g t g c g c g c t c c c g a c a t a t g c g a g a c t g n g c g a g a a g c t g n g a g	60
OY	211	c a g c t c g a g a t g n g a c a t a g n g a t g n g c g c g t g n g c g c g c g c g c g c c c c g a c g a c a g c	270
Db	61	c a g c t c g a g a t g n g a c a t a g n g a t g n g c g c g t g n g c g c g c g c g c g c c c c g a c g a c a g c	120
OY	271	t t c g c c a c c c a c c c t c g c c a g n a c a c c g t g a c t a t a c a a c c c a c c g a c t g t c g t t g g	330
Db	121	t t c g c c a c c c c c c t c g c c a c g n a c a c c g t g a c t a t a c a a c c c a c c g a c c t g t c g t t t g g	180
OY	331	g t c g a g a g c a t g c g t c g n g a g c t c a a c g c g c g c g c g c c c t c c c g c c g c c g c c a g	390
Db	181	g t c g a g a g c a t g c g t c g n g a g c t c a a c g c g c g c g c g c c c t c c c g c c g c c g c c a g	240
OY	391	c t c a a g c t c c a c c t c c t c c c a c c g t a c a g g c a g c g g c g a c t a c t c g a t c t c c g c c c	450
Db	202	c t c a a g c t c c a c c t c c t c c a c c g t a c a g g c a g c g g c g a c t a c t c g a t c t c c g c c c	261
OY	451	t c c g t c g a c t c t c c a g a g a t a t t a a g c g t g n g c g a g t c c c c t c c c c g n g c g g c g	510
Db	262	t c c g t c g a c t c t c c a g a g a t a t t a a g c g t g n g c g a g t c c c c t c c c c g n g c g g c g	321
OY	511	a c g c g c c g c g c a c g a c t c c g c c g a g a t c c g t g n g a g a t c c c a a c g n a t g c g a c t g n g c	570
Db	322	a c g c g c c g c g c a c g a c t c c g c c g a g a t c c g t g n g a g a t c c c a a c g n a t g c g a c t g n g c	381
OY	571	g g a g a c a g a c a c t c g t c g t a c t c t c c t c g t c	605
Db	382	g g a g a c a g a c a c t c g t c g t a c t c t c c t c g t c	416

XX	RESULT 14
XX	AA336269
XX	AA336269 standard; DNA: 436 BP.
XX	AA336269;
XX	
XX	16-JUL-1999 (first entry)
XX	
XX	DNA sequence obtained after sequencing wheat Rht clone 5a1.
XX	
XX	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
XX	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
XX	paclobutrazol; ss.
XX	
XX	Triticum aestivum.
XX	
XX	W09909174-A1.
XX	
XX	25-FEB-1999.
XX	
XX	07-AUG-1998; 98MO-GB02383.
XX	
XX	13-AUG-1997; 97GB-0017192.
XX	
XX	(PLAN-) PLANT BIOSCIENCE LTD.
XX	
XX	Harberd NP, Peng J, Richards DE;
XX	
XX	WPI; 1999-181040/15.
XX	
XX	New Triticum Aestivum polynucleotides - encode a polypeptide which
XX	provides inhibition of the growth of plants, which inhibition is
XX	antagonised by gibberellin, used to confer a dwarf phenotype
XX	
XX	Disclosure: Fig 2c(2); 88bp; English.
XX	
XX	The specification describes polypeptides encoded by the Rht gene (and
XX	its homologues) that, when expressed in Triticum Aestivum, inhibit
XX	growth of the plant. This growth inhibition is antagonised by
XX	gibberellin. The products can be used to provide Rht expression in
XX	





```
Db 1010 gaattggtccaccggaacccggaataattcgaatcattcgaatgagttgggtgtaagctgg 1069
QY 1202 ccgaattcgcgcacacacacacacacacacacacacacacacacacacacacacacacac 1261
Db 1070 ccaatttaagctgaagcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatc 1129
QY 1262 tcgcgcgaacccggaacccggaacccggaacccggaacccggaacccggaacccggaacccgga 1321
Db 1130 tagctgatacttgatcttgatcttgatcttgatcttgatcttgatcttgatcttgatcttgat 1174
QY 1322 ccgagagtaatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1381
Db 1175 ttgaaactctgttcggttaactctgttcggttaactctgttcggttaactctgttcggttaact 1234
QY 1382 cccctggagagaggttcctgggacacggttcggttcggttcggttcggttcggttcggttcggt 1441
Db 1235 cgatcgaataaggttccttggttggttggttggttggttggttggttggttggttggttggt 1294
QY 1442 agcagagagcgaatcacaacacacacacacacacacacacacacacacacacacacacacac 1501
Db 1295 agcagagaaacgaacacataatagtcgaatttcttagatcgtttactgagtcgttgcat 1354
QY 1502 actactcaacatgltcgaatccctcgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1561
Db 1355 attactcgaactgttggttggttggttggttggttggttggttggttggttggttggttggt 1385
QY 1562 tctcatcggggggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1621
Db 1386 -----tgtaacgagtggtcaagaacaaggtcaatgctcgagg 1420
QY 1622 tgtacctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1681
Db 1421 tttaacttggttaaacgaatcgtgcaacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1480
QY 1682 gccacgagacgctgggccaagtggcggaacccgctgggccaacccggttcgaagaccgtcc 1741
Db 1481 gtcatgaaacgltgagtcagtgaggaacccggttcggttcggttcggttcggttcggttcg 1540
QY 1742 acctgggctccaatgctcctacaagcagcgacgctgctgctgctgctgctgctgctgctgctg 1801
Db 1541 atattggttcgaatcggttaagcgaagcgaatcgtcttcttgctcctcgtcctcaacgagcggt 1600
QY 1802 acggtctacaaggctggaggaagaagagcgtgcctgacgctggg 1844
Db 1601 aggttatcgcggtgaggaagagtgacggtgtgtcatgtgtggg 1643
```

Search completed: January 23, 2002, 02:37:26  
Job time: 5950 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2002, 23:01:56 ; Search time 1805.74 Seconds  
(without alignments)  
12645.663 Million cell updates/sec

Title: US-09-485-529-14  
Perfect score: 2125  
Sequence: 1 atagagagcgcagtagctc.....tgtagtgacgacgactccg 2125

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapect 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_estin:\*  
3: em\_estom:\*  
4: em\_estpl:\*  
5: em\_estba:\*  
6: em\_estro:\*  
7: em\_estov:\*  
8: em\_hc:\*  
9: em\_hc1:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_hc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	563.8	26.5	778	11	BG308975 HVSMC000
2	516.4	24.3	597	11	BF620181 HVSMC001
3	503	23.7	851	11	BF254247 HVSMC001
4	464.2	21.8	563	11	BF588097 FMI_37_F0
5	445	20.9	593	10	BF588097 FMI_37_F0
6	421.2	19.8	551	10	BF588097 FMI_37_F0
7	382.2	18.0	443	11	BF588097 FMI_37_F0
8	381.4	17.9	559	11	BF618807 HVSMC000
9	365.2	17.2	440	11	BF618807 HVSMC000
10	335.2	15.8	400	10	BE414891 MML001.H0
11	329	15.5	481	11	BE417150 943053608
12	309.2	14.6	844	11	BF268018 HV_CEA001

13	307.4	14.5	527	10	BE595338
14	284.2	13.4	596	10	AU065169
15	273.4	12.9	524	10	BE329503
16	263.8	12.4	563	11	BG043071
17	258.8	12.2	543	11	BG155663
18	257.4	12.1	579	10	AW704479
19	256.2	12.1	467	10	AU222715
20	252	11.9	475	10	BE364489
21	251.4	11.8	658	11	BF587541 FMI_37_F0
22	244.2	11.5	588	10	A1665907
23	243	11.4	577	10	BE196286
24	242.2	11.4	568	11	BF201178
25	224	10.5	981	10	BE454560
26	223.4	10.5	743	10	AW774515
27	220	10.4	450	10	AW396192
28	219.8	10.3	626	11	BF009011
29	217.6	10.2	592	11	BG13881
30	215.6	10.1	494	11	BG510374
31	211.8	10.0	408	10	AW423888
32	211.8	10.0	714	10	AW348855
33	211.2	9.9	511	11	BG790472
34	209	9.8	416	11	BG051785
35	208.4	9.8	379	10	AV420689
36	207.2	9.8	452	10	A1416823
37	206.6	9.7	430	10	AV413998
38	201.4	9.5	700	11	BG587404
39	196.4	9.2	587	11	BE805895
40	190.2	9.0	473	10	AW734476
41	186.2	8.8	700	10	AL506960
42	182	8.6	626	10	BE204113
43	180.6	8.5	1064	10	BE035220
44	177.2	8.3	420	10	BE610936
45	173.4	8.3	671	11	BG443698

## ALIGNMENTS

RESULT 1	BG308975	778 bp	EST	22-FEB-2001
LOCUS	HVSMC0001C18f	Hordeum vulgare seedling shoot	EST library	
DEFINITION	HVCNMA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone			
ACCESSION	BG308975			
VERSION	BG308975.1	GI:13109822		
KEYWORDS	EST.			
SOURCE	Hordeum vulgare			
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.			
REFERENCE	Wing, R., Close, T.J., Kleinbofs, A., Wise, R., Begum, D., Frisch, D., Yu, Y., Anderson, H., Dale, J., Henry, D., Kennodie, S., Palmer, M., Ranbo, T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and Wood, T.			
AUTHORS	Development of a genetically and physically anchored EST resource for barley genomics			
TITLE	Unpublished (2000)			
JOURNAL	Clemson University Genomics Institute			
COMMENT	100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Seg primer: AATTAACCTCACTAAGCG High quality sequence stop: 690. Location/Qualifiers 1..778 /organism="Hordeum vulgare" /cultivar="Morex"			
FEATURES	source			

```

/db_xref="taxon:4513"
/clone="HVSMEC0018M1f"
/clone_lib="Hordeum vulgare seedling shoot EST library
HVCNMA0003 (Etiolated and unstressed)"
/tissue_type="Seedling shoot"
/lab_host="TUC121"
/note="Vector: lambdaZAP; Site.1: EcoRI; Site.2: XhoI. For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/ To
order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT      96 a      295 c      261 g      124 t      2 others
ORIGIN

```

```

Query Match      26.5%: Score 563.8; DB 11; Length 778;
Best Local Similarity 87.2%; Pred. No. 2.4e-68;
Matches 695; Conservative 0; Mismatches 74; Indels 28; Gaps 6;

QY 154 gccctcgagtaacagtgctgcgcctccagacatgagcagtgagcagagcag 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GCGCGGGGCTGCGAGAGGGGCGGCTCCGACATGGCGGAGCTGGCGCAGAGCTGAGCAG 60

QY 214 ctcgatagtcgcatgagtgagtgagtgagtgagtgagtgagtgagtgagtc 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CTCGAGATGGCATGGCATGGATGG-----CGGCCCGCCCGCGAGACGGCTTC 108

QY 274 gccacccaccctcgccagcagacgctgctactacaacccaccagacgtgctgtgctg 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 GCGACCCACCTTCGCGACGACGACCTCCACATCAACCCACCGACCTCTCTCGCTGGGG 168

QY 334 gagagatgctgctgagtgagtgagtgagtgagtgagtgagtgagtgagtc 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 GAGAGCATGCTGTCGAGCATCAACCGCGCGCGCGCGCTCCCGCGCGCGCGCAG 228

QY 391 ctcaagcgtccacgtctccacgttcagcagcagcagcagcagcagcagcagtc 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 CTCACCGCTCCACTCTCTCCACGCTCAAGGGGGGGGAGTACTTGCATCTCCGCCC 288

QY 451 tccgtgcgtctctccagacagatcagcagcagcagcagcagcagcagcagcagc 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 TCTGTGCACTCTCTCCAGACGACCTACGCTCGCCCGCGATCATCTCCCGCGCTC--- 345

QY 511 acgagccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 ---GCGCGCGCGCGACCTCTCGCTGATCTCGTCCGAGACCCCAACGGATGGCGACTGGC 402

QY 571 gggagcagacactgctgctatctctctctctctctctctctctctctctct 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 403 GGCAGCACACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 462

QY 628 agctctgtgtgagagcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 463 AGCTCTGTGTGAGAGGCTGTCTCCGCGGTGGGGCTGCGGC---TGTGCGCGCGCGCTG 519

QY 688 ccggtctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 CCGGTCTGTGTGAGAGGCTGTCTCCGCGGTGGGGCTGCGGC---TGTGCGCGCGCGCTG 579

QY 748 tgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 580 TGCgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 639

QY 808 cccctgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 640 CCCCTGTGTGAGAGGCTGTCTCCGCGGTGGGGCTGCGGC---TGTGCGCGCGCGCTG 699

QY 868 gccctcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 927
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 700 GCGCTTGGCGCGCGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 758

QY 928 gccctcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 944
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 759 GCGCTTGTGTGAGAGCTTCT 775

```

```

RESULT 2
BF620181
LOCUS
DEFINITION
    BF620181
    HVSMEC0018M1f Hordeum vulgare seedling shoot EST library
    HVCNMA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone
    HVSMEC0018M1f, mRNA sequence.
ACCESSION
    BF620181
VERSION
    BF620181.2 GI:13109232
KEYWORDS
    EST.
SOURCE
    barley.
    Hordeum vulgare
    Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
    1 (bases 1 to 597)
    Triticaceae; Hordeum.
REFERENCE
    1 (bases 1 to 597)
    Wing,R., Close,T.J., Kleinbols,A., Wise,R., Begum,D., Fritsch,D., Yu
    ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
    ,T., Saski,C., Schwartzbeck,J., Simons,J., Chol,D.W., Main,D. and
    Wood,T.
    Development of a genetically and physically anchored EST resource
    for barley genomics
    Unpublished (2000)
    On Dec 18, 2000 this sequence version replaced gi:11883915.
COMMENT
    Contact: Wing RA
    Clemson University Genomics Institute
    Clemson University
    100 Jordan Hall, Clemson, SC 29634, USA
    Tel: 864 656 7288
    Fax: 864 656 4293
    Email: rwing@clemson.edu
    Seq primer: AATTAAACCTCCTCACTAAAGG
    High quality sequence stop: 564.
    Location/Qualifiers
        1..597
            /organism="Hordeum vulgare"
            /cultivar="Morex"
            /db_xref="taxon:4513"
            /clone="HVSMEC0018M1f"
            /clone_lib="Hordeum vulgare seedling shoot EST library
            HVCNMA0003 (Etiolated and unstressed)"
            /tissue_type="Seedling shoot"
            /lab_host="TUC121"
            /note="Vector: lambdaZAP; Site.1: EcoRI; Site.2: XhoI. For
            more details on library preparation and sequence analysis
            see http://www.genome.clemson.edu/projects/barley/ To
            order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT      99 a      202 c      204 g      91 t      1 others
ORIGIN

```

```

Query Match      24.3%: Score 516.4; DB 11; Length 597;
Best Local Similarity 92.5%; Pred. No. 7.3e-62;
Matches 554; Conservative 0; Mismatches 42; Indels 3; Gaps 1;

QY 1298 ggcaggaagacccgagagagagcagcagcagcagcagcagcagcagcagcagc 1357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GCGAGGAGGACCTTAACAGAGGCCGAGGATATGCGGTGATCACTTCTGAGATGC 60

QY 1358 accgctgctcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 ACCGCTCTCTCGCGAGCGCGCGCGCTGTGAGAGCTCTGTGGCGAGGTCGCGCGCTGC 120

QY 1418 gccacagatcgtaacgctgtgtgagagagagcagcagcagcagcagcagcagc 1477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GCGCATATGCTACACCTGTGTGAGAGGAGGCGAGGACCAACACTCCGCTCATTCCTGG 180

QY 1478 accgcttaccagagctgtgactactactactactactactactactactactact 1537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 ACCGCTTACCGAGTCTCTGCACTACTACTACCATATTCGATTCCTCTCGAGGGGCA 240

QY 1538 gtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1597

```

```

Db      241 GCTC---CGGGCGCCCGCTCGAGGTCTCATCGGGGGGCTCCCTCTCTGCGCGCGCGCG 297
QY      1598 gaacgaccaggtatcgttcgagtgtagcttcgagcgagatctgcaagctgtgtgct 1657
Db      298 GCACGGACACGAGTCAATGTCGAGGTCTGACCTGCGCGGAGATCTTCAACGCTGTGGCT 357
QY      1658 ggaagagggcgagagcgacagagcgccacagagcgctgagcgagtggtgagagcgagctg 1717
Db      358 GCGAGGGACGAGGAGCGACAGAGCGGACGAGACACTGAGGAGTGCGGAGACCGGTTGG 417
QY      1718 ggaacgaggttcgagagcgatccacgctgagctccatctaccacagcgagcgagcgacg 1777
Db      418 GCAAGCGCGGGTTGAGAGCGTGGCACTGGGCTCCATCTCAATCAAGCAGCGAGACCG 477
QY      1778 tgcctgagcttcgagcgagcgagcgagctacagagtgagagagagagagctgctga 1837
Db      478 TCTCGGCTCTCTTCGCGCGCGCGCGAGCTGACAGGTGAGAGATGGAAGGTCGCTGA 537
QY      1838 gctgaggtgagacagcgccgctgagcgacgctgagcgagcgagcgagcgagcgag 1896
Db      538 CTCTCGGGTGGACACGCGCCGCTGATGCGCACTTCCGATGGCGCTCGCGCGCGCG 596

```

```

RESULT 3
LOCUS   BF254247      851 bp      mRNA      EST      23-FEB-2001
DEFINITION
HVSMEF0003G24f Hordeum vulgare seedling root EST library HVCDA0007
(etiolated and unstressed) Hordeum vulgare cDNA clone
HVSMEF0003G24f, mRNA sequence.

```

```

ACCESSION BF254247
VERSION    BF254247.2
KEYWORDS   GI:13117026
SOURCE     barley.

```

```

ORGANISM   Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
            ; Triticeae; Hordeum.
            1 (bases 1 to 851)
            Wing,R., Close,T.J., Kleinof,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kennodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Chol,D.W., Main,D. and
            Wood,T.
            Development of a genetically and physically anchored EST resource
            for barley genomics
            Unpublished (2000)
            On Nov 16, 2000 this sequence version replaced gi:11183352.

```

```

JOURNAL    Contact: Wang RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rtwing@clemson.edu

```

```

COMMENT    Seq primer: AATTAACTCTCACTAAAGG
            High quality sequence start: 27
            High quality sequence stop: 611.
            Location/Qualifiers
            1..851
            /organism="Hordeum vulgare"
            /cultivar="Morex"
            /db_xref="taxon:4513"
            /clone="HVSMEF0003G24f"
            /clone_1db="Hordeum vulgare seedling root EST library
            HVCDA0007 (etiolated and unstressed)"
            /tissue_type="Seedling root"
            /lab_host="TJUC121"
            /note="Vector: lambdaZAP. Site 1: EcoRI. Site 2: XhoI. For
            more details on library preparation and sequence analysis
            see http://www.genome.clemson.edu/projects/barley/ To
            order a clone see http://www.genome.clemson.edu/orders"

```

```

FEATURES
Source

```

```

BASE COUNT 151 a      273 c      288 g      139 t
ORIGIN

```

```

Query Match      23.7%; Score 503; DB 11; Length 851;
Best Local Similarity 82.4%; Pred. No. 5,1e-60;
Matches 642; Conservative 0; Mismatches 120; Indels 17; Gaps 5;

```

```

QY      1303 gaagaccgaacagagagcgacgagatcgcgcgttaactcaactcttgagatgaacgg 1362
Db      2 GAGAGACCCCCCGAGAGAGCCCGAGTAATCCCTTAACCTCACTTTCGATGACCCG 61
QY      1363 ctgctcgagcagcccgagcccttgagagagtgctcttgagacgctgagcgctgagcc 1422
Db      62 CTCCTCGCGAGCGCGCGCGCTCGAGAGGCTCTGGGACGAGTGGCGCGCGCGCGCG 121
QY      1423 aggatcgtacagctgtgtgagcagagagcgagatcaactcgcgcaattcttgagcgc 1482
Db      122 AGGATCTAAACCCCGGTGAGAGGAGGAGAACCAACATCCCGCTCATCTCGAGCCG 181
QY      1483 ttacagagctctctgactactactacacatgttgattccctgagagcgagctgc 1542
Db      182 TTCACGAGTCCCTGCACTACTACTCCACCATGCTGATTTCTTCAGAGGAGGAGCTC- 240
QY      1543 ggcgagcgccacalcgaaagctcactcgggggctgctgctgctcgcgcgcgcgcg 1602
Db      241 --CGAGGCGCCGTCGAGGCTCATCGGGGGGTGCCGTCCTCCCGCGAGCGCGGCGAG 298
QY      1603 gaccaggtatctcgagagtgtaactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1662
Db      299 GACCAAGTATGTCGACAGGTGTAACCTCGGCGCGCAATCTGCAACCTGGTGGCTTCGAG 358
QY      1663 ggggagagcgacagagagcgcccgagcgtctgagcgagtgagagcgagcgagcgagc 1722
Db      359 GGCATGAGCGACAGAGCGCGCCCGGAGACCTGGGCACTGGCGGAGACCGGCTGGGCAAC 418
QY      1723 gcccgttcgagacgctccacgttgctccaatgctccataagcagcgagcagcgtgctg 1782
Db      419 GCGGGTTGAGAGCGCTGACACTGGGCTCCAAATGCCCTAC-AGCAGCGAGACAGCTGCTG 477
QY      1783 ggcctcttcgagcgagcgagcgtacagagtgagagagagagagagcgtcgtcgaagcgc 1842
Db      478 GCCCTTTGGCGCGCGCGGACGCTGCTACGTGAGAGAGACAGAGGCGGCTGACTCTC 537
QY      1843 ggttgacacagcgcccgctgactgcacgtcgacgtcgagctgagcgagcgagcgagc 1902
Db      538 GGGTGACACAGCGCGCCGCTGATCGGCACCTCCGCAATGGCGGCTTGTGGCGGCGGA--T 595
QY      1903 cgcgagtttgaagcgtgtgaatgacacatcgtgag-----catggaagacacac 1952
Db      596 CGGAGTATGAAGCGCTGTGGAAAGACTTGTGAGAGCGTGAGCGCTACAACTCAACCC 655
QY      1953 agcccgagcgagcgcccgctctcgcggaagcagcagcagcagcagcagcagcagcag 2012
Db      656 CGGCGCGCTCCGCTCCGCGGTCTGGGGCCCGCCACCCCTTTAGAGCAAGAGAT 715
QY      2013 gaagcca-aatgcatgtaagtgagcgtgaattgagcgagcagcagcagcagcagcag 2070
Db      716 GGAATTAAGAACTAATGGCGGAGTCTCTGATGGGTGAACCGGAGATGATGAGCG 774

```

```

RESULT 4
LOCUS   BF588097      563 bp      mRNA      EST      12-DEC-2000
DEFINITION
FML_37_F08_b1.A003 Floral-Induced Meristem 1 (FML) Sorghum
PROPIONQUUM cDNA, mRNA sequence.
ACCESSION BF588097
VERSION    BF588097.1
KEYWORDS   GI:11680421
SOURCE     Sorghum propionquum.
ORGANISM   Sorghum propionquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 563)

```

AUTHORS	Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M., Marsala, C. and Pratt, L.H.
TITLE	An EST database from Sorghum: floral-induced meristems
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cordonnier-Pratt MM

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpatt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
Seq primer: JEN REV  
High quality sequence stop: 477  
POLYA-No.

FEATURES	Location/Qualifiers
source	1. .563

```

/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Floral-Induced Meristem 1 (FIM1)"
/note="Organ: Floral-Induced meristems; Vector:
pBluescript II from Lambda Zap II; Site.1: XhoI; Site.2:
EcoRI; mature plants were placed in a growth chamber for
15 days with 16 hr darkness and 8 hr light (flowering is
induced by short-day conditions); 16 days after being
returned to the greenhouse under natural long days during
late April/early May, meristems were harvested. The
library was made from poly-A RNA in the cloning vector
lambda Zap II. Clones to be sequenced were prepared by
mass excision."

```

BASE COUNT	76 a	217 c	175 g	95 t
ORIGIN				
Query Match	21.88;	Score 464.2;	DB 11;	Length 563;
Best Local Similarity	90.38;	Pred. No. 1e-54;		
Matches 496;	Conservative	0;	Mismatches 53;	Indels 0;
				Gaps 0;

[illegible]

Db	QY	Db
495	1218	555
GCACGACGAGCCACGCGCTTGCACAAGTGGCTGGAAACTTGTCTAGTTCCGGCACAC	catccgcgt	CATCCGCGT
	1226	563

RESULT	5
LOCUS	A1881894
DEFINITION	A1881894 593 bp mRNA EST 02-FEB-2000
VERSION	60607ACG8.Y1 606 - Ear tissue cDNA library from Schmidt lab Zeeland
KEYWORDS	mays cDNA, mRNA sequence.
SOURCE	A1881894.GI:5566983
	EST.
	Zea mays.

ORGANISM  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE  
1 (bases 1 to 593)

AUTHORS	Walbot,V.
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL	Unpublished (1999)
COMMENT	Contact: Walbot V

Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2237  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 606074 row: 5 column: 08.

```

FEATURES
    source
        location/Qualifiers
            1..593
                /organism="Zea mays"
                /cultivar="Oh1043"
                /db_xref="taxon:4577"
                /clone_1id="606 - Ear tissue cDNA library from schmidt
                    lab"
                /tissue_type="mixed"
                /dev_stage="ear length from 0.5 cm - 2.0 cm"
                /lab_host="X10LR (Stratagene)"
                /note="Organ: Immature ear; Vector: pBK-CMV; Site_1: EcorRI
                    ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                    lab"
BASE COUNT
    93 a    210 c    208 g    82 t
ORIGIN

```

Query Match	20.9%	Score 445:	DB 10:	Length 593:
Best Local Similarity	85.2%	Pred. No. 4.3e-52:		
Matches 512:	Conservative	0:	Mismatches 80:	Indels 9:
				Gaps 1:
QY 1226	tgcattccagttacacgcgcgtctgcgcgcgcacacgtctcgcgagacctggagccgttcatg	1285		
Db 1	TGCATTTCACGTAACGCTGTGCTCGCGCGCAGCTCTGCTATCCTGGAGCGCTTCATGC	60		
QY 1286	tgcacgcgagagggcaggaagaacccgaacagagagccgaggttaatcgccgttcaactcaag	1345		
Db 61	TGCGACCGGAGGGCGCACGCGCACAGGAGAGAGACCCCGAGGTGATCCCGTAAACTCGG	120		
QY 1346	tcttcgagatcacacgcgctctgcgcgcgcacccgcgcgccttgagaagaagttcctgggcacg	1405		
Db 121	TGTGCGAGCTGCACCGGCTGCTCGCGCACCCCGGTATCACTGCACAAAGTCTCTGGGCACCG	180		
QY 1406	tgcgcgcgctgcgcgcgcacagatctgtcacacgttgctggacagagagcgatcatcaactccg	1465		
Db 181	TGCGCGCGGTGCGGCGCGAGATCGATCGAGCGGTGTGAGACAGAGGCCAACCACAACTCCG	240		
QY 1466	gcaeatctctggaaccttcaacagagatctctgacactactcaactcaacatggttcgaattccc	1525		

[illegible][illegible]

RESULT	7
BF655617	
LOCUS	BF655617 443 bp mRNA EST 20-DEC-2000
DEFINITION	FM1_46.D10.b1.A003 floral-induced Meristem 1 (FM1) Sorghum
ACCESSION	propinquum cDNA, mRNA sequence.
VERSION	BF655617
KEYWORDS	BF655617.1 GI:11920749
SOURCE	EST.
ORGANISM	Sorghum propinquum.
REFERENCE	Sorghum propinquum
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC clade; Panicoidae; Andropogoneae; Sorghum. 1 (bases 1 to 443) Cordonier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt L.H.
TITLE	An EST database from Sorghum: floral-induced meristems
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cordonier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. Seq primer: JEN REV High quality sequence stop: 395 POLYA-No.







QY	1068	ggggatgcagttggcccccactttcttcaggccctctgcgcccctctccgagggcccttcctc	1127
Db	181	GGGGATCAGTGGCCGCCCTTCTCCAGGAGCTTGAGGCTCGCTCCCGGGGGCCCTCTCCTC	240
QY	1128	gttcgcctcaacggcgcgtcgccccccagccagccgcgcagacagacagccctgcagcaggt	1187
Db	241	GTTCGCG-CTCACCAGCGGCTGGCCCCCGCAGCGGAGCGAGACCGAGCGCTTGACAGCAAGT	299
QY	1188	ggcctgtgaagctcgcccaagttcgcgcaacacatccgcgtctgcattccagttacgcgcgct	1247
Db	300	GGGCTGGAGAACTGGCCCAAGTTCGCGGACACACATCCGCGCTTCGACTTCAGTACCGNGGCT	359
QY	1248	cgtcgcgcacacgcctcgcgagcactgtgagccgt	1279
Db	360	CGTTGC--GCACGCTTNGNGGACCTGGAGCGCGT	389
RESULT	11		
LOCUS	BI417150	481 bp	EST
DEFINITION	949053G08.y1_949	juvenile leaf and shoot cDNA from Steve Moose	15-AUG-2001
ACCESSION	BI417150	Zea mays cDNA, mRNA sequence.	
VERSION	BI417150.1	GI:15188173	
KEYWORDS	EST.		
SOURCE	Zea mays.		
ORGANISM	Zea mays.		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta::		
AUTHORS	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC		
TITLE	clade; Panicoideae; Andropogoneae; Zea.		
JOURNAL	1 (bases 1 to 481)		
COMMENT	Maise ESTs from various cDNA libraries sequenced at Stanford University		
COMMENT	Unpublished (1999)		
COMMENT	Contact: Walbot V		
COMMENT	Department of Biological Sciences		
COMMENT	Stanford University		
COMMENT	855 California Ave, Palo Alto, CA 94304, USA		
COMMENT	Tel: 650 723 8227		
COMMENT	Fax: 650 725 8221		
COMMENT	Email: walbot@stanford.edu		
COMMENT	Plate: 949053	row: G	column: 08.
FEATURES	Location/Qualifiers		
source	1. 481		
	/organism="Zea mays"		
	/cultivar="W64A"		
	/db_xref="taxon:4577"		
	/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve Moose"		
	/tissue_type="Immature leaf primordium and vegetative meristem"		
	/dev_stage="4 stages from 3-13 days after imbibing"		
	/lab_host="E. coli XL0LR"		
	/note="Organ: juvenile vegetative shoots; Vector: pMD-Gal4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts of total RNA by weight from 4 tissue sources (see below) were pooled, polyA+ RNA isolated, and cDNA synthesized for EcoRI (5') and XhoI (3') directional cloning into lambda HybriLap vector from Stratagene. Tissue Sources: 1. Whole shoots 3 days after sowing/imbibing in wet soil. 2. Basal 1.5 cm shoots 6 days after sowing - includes yellow portions of developing leaves 1-5, primordia from 6-8, and the vegetative apex. 3. Non-green portions of developing leaves 4-5 and the vegetative apex, including adult leaf primordia, 9 days after sowing. 4. Partially expanded and greenling leaves 4-5 at 13 days after sowing."		
BASE COUNT	62 a	185 c	150 g
ORIGIN		75 t	9 others
Query Match	15.5%	Score 329;	DB 11; Length 481;

[illegible]

```

/clone="HV_CEA0019J17f"
library HVCDNA0004 (Erysiphe infected & control)
/tissue_type="seedling green leaf"
/lab_host="TJ121"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/To
order a clone see http://www.genome.clemson.edu/orders"

```

```

BASE COUNT      150 a      290 c      276 g      128 t
ORIGIN

```

```

Query Match      14.6%; Score 309.2; DB 11; Length 844;
Best Local Similarity 79.5%; Pred. No. 1.7e-33;
Matches 410; Conservative 0; Mismatches 88; Indels 18; Gaps 3;

```

```

QY 20 cgcgcatcgtgaagcggagatcaccagacgcggagggagcgcgctgagcgca 79
DB 138 CGAGATCTGTAAGCCGACGACGACGAGGCGGCGGCGGCGGTGGGTGATGAGA 197
QY 80 tgggctctccagaggacaagatgagt---gtcgcgcgcgcgcgaggagggagag 136
DB 198 TGGGGTCGTGAGGGGCAAGATGATGTCGTGCGCCGAGGCGGGGAGAGAGAG 257
QY 137 tggagagctgtcgtcgcgcgctcggtacaaagtgcgcctccagatgagcgatg 196
DB 258 TGGAGAGCTGCTGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 317
QY 197 cgcagaagctgtagagagctcgagatgagatggagatggcgcgcgcgcgcgcg 256
DB 318 CCGAGAGCTGAGAGAGCTCGAGATGCGCATGAGATGGG-----CGGGCCTT 365
QY 257 ccccgagacagacgtctcgccaccacgtcgccagagacagctgacatacaaccacg 316
DB 366 CCGCGAGACGCGCTGTGACCCACCTGACGACGAGACGCGGCGCATCAACCCACG 425
QY 317 acctgtctcttgggtcgagagatgctgtcgagatcaagcgcgcgcgccctcc 376
DB 426 ACCTCTCTCTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485
QY 377 ggcgcgcgcgc---cgagagctcaagcgtcctcctcctcctcctcctcctcctc 433
DB 486 CCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 545
QY 434 acctgactccgcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 493
DB 546 ACTTATGATCTCCGGGCTCTGTCGACTACTCCAGCAGAAACTAGCCCTGGGGCCCGAT 605
QY 494 cctcccgcgcgcgcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 529
DB 606 CATTCCTCCGCGACCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 641

```

RESULT 13

```

LOCUS      BE595338      527 bp      mRNA      EST      18-AUG-2000
DEFINITION P11_48-G12.b1-A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
ACCESSION BE595338
VERSION BE595338.1 GI:9850411
KEYWORDS EST.
SOURCE sorghum, bicolor
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 527)
Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
L.H.

```

```

TITLE      An EST database from Sorghum: pathogen-induced plants
JOURNAL    Unpublished (2000)
COMMENT    Contact: Cordonnier-Pratt MM

```

```

Department of Botany
The University of Georgia
Plant Sciences Building, Km. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmprratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 524
POLYA-No.

```

```

FEATURES
Source
Location/Qualifiers
1..527
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen induced 1 (P11)"
/note="Organ: Anthracnose-infected leaves from
two-week-old sorghum plants 48 hr after inoculation;
vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
cultivar) were infected with pathogen (isolate PRM421 of
Colletotrichum graminicola, which is a sorghum isolate).
RNA was prepared from infected leaves harvested from 45
seedlings 48 hours after inoculation. Note: young
seedlings (2 weeks old) exhibit juvenile resistant
reaction, which is an incompatible interaction. As they
grow older (4 weeks or older), plants resume susceptibility
to anthracnose disease. The library was made from poly-A
RNA in the cloning vector lambda Zap II. Clones to be
sequenced were prepared by mass excision. WARNING: While
most or all ESTs are expected to derive from the host
plant, no effort was made to eliminate ESTs deriving from
the pathogen."

```

```

BASE COUNT      65 a      197 c      183 g      82 t
ORIGIN

```

```

Query Match      14.5%; Score 307.4; DB 10; Length 527;
Best Local Similarity 78.3%; Pred. No. 2.9e-33;
Matches 397; Conservative 0; Mismatches 101; Indels 9; Gaps 2;

```

```

QY 398 cctccaccctccaccgctcagcgagcgcgctcctcctcctcctcctcctcct 457
DB 21 CCACCTGACAAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80
QY 458 actcctcagacatcagcgtcgcgcgatccctcccgcgcgcgcgcgcgcgcg 517
DB 81 ACTCGTCCAGCAGCACTTACGCTTAAAGCCATCCCTGCGCGGTGGCGCTCGG 140
QY 518 cgcgcgacctgtcgcgcgacatcgcgtgagatcccaagcgatgacgagcgag 577
DB 141 CCGACCGCTCTCCAGCAGCACTCCAGCGGAGCGCAAGCGGAGTGGCGCGCGCA 200
QY 578 gaaacctgtctatctcctcctcctcctcctcctcctcctcctcctcctcct 634
DB 201 GCAGCTGCTTCTGCTTCTCTCTGTCATCATGATGATGGGCGCGCGCGCGCG 260
QY 635 tggtagagctgcggcgcgctc-----gcgcgcgcgcgcgcgcgcgcgcgcgc 688
DB 261 TGGTCAAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 320
QY 689 cgtctgtgtgtcgaacgagcgagggatcgtcgtgtgtcgaacgagcgctgtg 748
DB 321 CGGTGTGTGTGATGACAGCAGAGAGAGTGGATCGGCTGTGTCACGGCTGTGCG 380
QY 749 ggcgagagcgcgcgagcgaggaactcctcgcgcgcgcgcgcgcgcgcgcgcgc 808
DB 381 GCGCGGAGCGCGCGAGCAGAGAACTTCTTCTGCGCGAGCGCGGTGTCAGCAATCC 440
QY 809 ccttgtagcgcgctcccgagcgcgctgagatgcaaggtcgcgcgctcctcctc 868
DB 441 CCATGTGGCTGTGTGCGAGGCGGTGCGCATGGCAAGTGGCGCGCTACTTGGCGAG 500

```

[illegible]

Db	300	gcccattcttccttacccgggtggtggccagc---cttaccctctgcctgcctgaatcgtggcag	356
QY	546	ggatcccaagcggatgcgcaactgtgcgggaagcaacccttcgtcatcctcctcgtc	605
Db	357	ggacaccaaagcgatgacccactgacggcgagcagcagtcgctcctcctccttc	416
QY	606	gtctct---cgggtgagggcgcacgaagctctgtgtgaaagctgc-----cgscgtc	657
Db	417	ctctctggcggctggggcctctgcggggctctgtgtgtgagctgtgtcccgccgacgca	476
QY	658	gcgcgcgcgcgcacaaagcgcacgcgcgcgtctgcgtctgtgtgtgcacagcagaagcc	717
Db	477	ggggccggcgccggcgcgaatgcgcgcgcg---tgccgggtgtgtgtgtgtgacacgagagct	535
QY	718	gggattgcgtgtgtgc-acgcgctgtctgtgcgtgtgcggaagcgcgtgcacgaagaacct	776
Db	536	gggatccgcgctgctgcacacgcgcttcctgcgtgcgcgcgagccgctgcacaggaacct	595
QY	777	c 777	
Db	596	c 596	
RESULT	15		
LOCUS	BE329503	524 bp	EST 31-JUL-2001
DEFINITION	so66b12.y1 Gm-cl040 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl040-600 5' similar to TR:065367 065367 RGAL.GENE.; mRNA sequence.		
ACCESSION	BE329503		
VERSION	BE329503.1	GI:9203279	
KEYWORDS	EST.		
SOURCE	soybean.		
ORGANISM	Glycine max		
REFERENCE	Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.		
AUTHORS	1 (bases 1 to 524)		
TITLE	Shoemaker,R., Keim,P., Vodkin,L., Erpelidng,J., Corvelli,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucab,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harrey,N., Schurk,R., Riltter,E., Kohn,S., Shin,T., Jackson,T., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.		
JOURNAL	Public Soybean EST Project		
COMMENT	Unpublished (1999) Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available through: Genome Systems, Inc. 463 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com Insert Length: 1533 Std Error: 0.00 High quality sequence stop: 457. Location/Qualifiers 1..524 /organism="Glycine max" /db_xref="taxon:3847" /clone="GENOME SYSTEMS CLONE ID: Gm-cl040-600" /clone_id="Gm-cl040" /tissue_type="Hypocotyl and plumule, germinating seeds" /lab_host="DH10B" /note="Vector: pRT3Pac (Pharmacia); Site 1: EcoRI; Site 2: NotI; This cDNA library was constructed from mRNA isolated from hypocotyl and plumule tissues of seeds		







[illegible]

RECORD 2  
 US-09-117-853-3  
 : Sequence 3, Application US/091178533  
 : Patent No. 6307126  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Harberd, Nicholas P  
 :  
 : APPLICANT: Peng, Jintong  
 : APPLICANT: Carol, Pierre

```

APPLICANT: Richards, Donald E
TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
FILE REFERENCE: 620-45
CURRENT APPLICATION NUMBER: US/09/117,853
CURRENT FILING DATE: 1998-08-12
EARLIER APPLICATION NUMBER: PCT/GB97/00390
EARLIER FILING DATE: 1997-02-12
EARLIER APPLICATION NUMBER: GB 9602796.6
EARLIER FILING DATE: 1996-02-12
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1643
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-117-853-3

```

Query Match	16.6%;	Score 351.8;	DB 4;	Length 1643;
Best Local Similarity	59.3%;	Pred. No. 3.8e-50;		
Matches 701; Conservative	0;	Mismatches 407;	Indels 75;	Gaps 32;

[illegible]



[illegible]

```

; RESULT: 3
; US-09-117-853-9
; Sequence 9, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jinrong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117, 853
; CURRENT FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: PCT/GB97/00390
; EARLIER FILING DATE: 1997-02-12
; EARLIER APPLICATION NUMBER: GB 9602796.6
; EARLIER FILING DATE: 1996-02-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1642
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-117-853-9

```

Query Match	16.3%;	Score 347;	DB 4;	Length 1642;
Best Local Similarity	59.68;	Pred. No. 2.4e-49;		
Matches 686; Conservative	0;	Mismatches 390;	Indels 75;	Gaps 3;

[illegible]

Dd	747	g c g c g g a g n a t t a c c g t c t c t c t c g t c g a g ----- a g t c a a t c g a c c a c t c t c	800
Qy	934	g c c g a c c c t c c a c g c g c a c c t t c a g a t t c c t g c c c t a c a a g t t c g c g a c t c	993
Dd	801	t c g a t a c t c t c a a g a t a g a c a c t t c a a g a g a c t t g c t a t c t a a g t t c g t c a a c t c	860
Qy	994	a c c g c c a a c c a a g g c a c t c c t g a g a g c g t t c c g c g t c g c c g c g c y g a a g t c g t a c	1053
Dd	861	a c c g c g a t c a a a g a t t c t c g a a g c t t t t c a a g g g a a a a a g a g t t a t g t a t g a t	920
Qy	1054	t t c g c a t c a a g a g a g g a g a g a t g c c c a c c t t c c a g a c c t c a g a c c t c g c c t c g t c c	1113
Dd	921	t t c t c t a t a g t a a a g t c t c t a a t a g g c g g c t t a g a g c g t c t t g c g t t c g a c t	980
Qy	1114	g g c g c c c t c c t g t t c c g c t c a c c a c c g c g t c g c c c c g c a g c g g a c g a a c g a c	1173
Dd	981	g g t g c t c c t c t g t t t c t c g t t a a c c g a t t g t c a a c c g a c c g a c c g a t a a t t c g a t	1040
Qy	1174	g c c c t t g a a g a g g t g g g c t t g a a g a g c c g a g t t c g g a c a a c a t c c g t c g a t t c	1233
Dd	1041	t a t c t a c t a a g a t t g g g t g t a a a g c t g g c t c a t t a a g c t a a g a g a t t a a c g t t a a g t t	1100
Qy	1234	a g t a a c c g c g a c c c t g c g c a c c a c a c g t c c g a a c c t t g a a g c g t t c a t g t c g a a c g	1293
Dd	1101	a g a t a c a a g a a t t t g t g g t c a a c c t t a a g t a g c t g a t c t t g a t g t c g a t c g a t g t t a g c -	1158
Qy	1294	a a g a g c g a a g a a c c c g a a c a g a g a g a c c c g a g t a t a t c g c g t c a a t c a g t c t t g a g	1353
Dd	1159	----- t t a a g a c a a a g t a g a t t a g a t c t g t c g t a a c t t c t t t c g a g	1205
Qy	1354	a t g a c c c g c t g c t g c g a c c c g c c c c t g a a a a g t c c t t g g a c c g t t g c g c	1413
Dd	1206	c t t c a a a g c t c t t g g a g a c a c t g t t g c a t c g a t a a a g t t c t g g t g t g t a a a g	1265
Qy	1414	g t g c g g c c c a a g a t c g t a c a c c t g t g a g a a g a g a g a t c a a t c a a c t c c g a c a t c	1473
Dd	1266	a t t a a a c c g a t t t t c a c t t g t t g t t g a g a a g a a t c g a a c a t a a t a g t c g a t t t c	1325
Qy	1474	c t g a a c c g t c a a c c a g a t c t t c g a c t a a c t a a t c a a a t t t g a t a c c t c g a a g g c	1533
Dd	1326	t t a a a t c g t t a c t a g a t c g t a g c a t t a t a c t a a c g a t g t t g a c g t c g t t g a a a g -	1384
Qy	1534	g a c a g c t c c g c g c g c c a t c a c a g t t c a t c g g g g t c g t c t c t c t c c g c c	1593
Dd	1385	----- t g t a c g	1391
Qy	1594	g c c g c a a c g a c a g a t c a g t c c a g a g t a c t c g c c g a c a g a t c t g a a c g t g t	1653
Dd	1392	a g t g t c a a a a a g a g t c a t g t c g a g a g t t t a c t t g g t a a a c a a g a t c t g a a c g t g t	1451
Qy	1654	g c c t g c a g a g g a g a g a c a a a g a c g a c g a a a a c a c t t g g c a a g t g g a g a a c c g	1713
Dd	1452	g c t t g t a t a g a c c t a g c a g a g t t g a g c g t a g a a a c g t t a g t c a g t g a g g a c c g	1511
Qy	1714	c t g a g c a a c c g a g t t c g a a a c c g t c a a c c t g g c t c c a a t a g c t a a a g a g a g a g c	1773
Dd	1512	t t c g g t c t g c t g g g t t t g c g t c g a c a t a t g t g t c t a a g a t a g c g t t a a g a a a g c g a t	1571
Qy	1774	a a c g c a g a g a g c c t t c c g a c c g a g a a g g t a a a a g t t g a a g a a a a a a a g a a c g c g	1833
Dd	1572	a t g c t t t t g t c t c g t c a a c c g c g t a g a g g t a t c a c g t g t g a g a g a g t g a a c g c g t	1631
Qy	1834	c t a a c g c t g g g	1844
Dd	1632	c t a a t t t g g	1642

RESULT 4  
US-09-117-853-5  
; Sequence 5, Application US/09117853  
; Patent No. 6307126  
; GENERAL INFORMATION:





[illegible]

Db	24213	CCCCACCAAGGCGATGCGC-----GAGCCCTTGACCGCGCATGCTGAGGCGGCGACACGCGTGG	24268
Qy	1547	gagagccaatcgaaatctcatatgtagggagctgcgtcgtctctctgacgagcgagcaacgagac	1606
Db	24269	GTCGCGTGGCGGACACTTCGATGAGTGAACCGGTTGGCCGCTTCACACTTGAGTGGTCCG	24328
Qy	1607	aggtcatgtcagaggtgtactctcgccgagcag	1638
Db	24329	CCGTACTCTTCGACGACATCCCGAGCGTGCAG	24360

RESULT 7  
US-08-804-198-1

; Sequence 1, Application US/08804198  
; Patent No. 5945320

```

; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.

```

APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rao, Nagaraja R.

APPLICANT: Richardson, Mark A.  
APPLICANT: Rosteck, Paul R., Jr

```

; TITLE OF INVENTION: PLATENOLIDE
; NUMBER OF SEQUENCES: 6

```

;; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PAUL R. CANTRELL

```

;
; STREET: LILLY CORPORATE CENTE
; CITY: INDIANAPOLIS

```

```

; STATE: IN
; COUNTRY: USA
;

```

```

; ZIP: 46285
; COMPUTER READABLE FORM:
;

```

```

; MEDIUM TYPE: floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh

```

```

; OPERATING SYSTEM: Macintosh /
; SOFTWARE: Microsoft Word 5.1

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804

```

```

; FILING DATE:
; CLASSIFICATION: 435
;

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
;

```

REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P911

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3885

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 44377 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: DNA (genomic)
; FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 350..14002
;

```

```

; FEATURE: CDS
; NAME/KEY: CDS
; POSITION: 14010 20000

```

```

; LOCATION: 14046..20036
; FEATURE:
; VALUE/UNIT: 000

```

```

; NAME/KEY: CDS
; LOCATION: 20110..31284
;

```

	FEATURE:	CDS
; ;	NAME/KEY:	
; ;	NUMBER:	01000 00071

```

; LOCATION: 31329..360/L
; FEATURE:
; VALUE: 000

```

```

; NAME/KEY: CDS
; LOCATION: 36155..41830
;
; 00 004 100 1

```

05-08-804-198-1

Query Match	4.78;	Score
Query Match	4.78;	Score

```
Best Local Similarity 45.48; Pre
Matches 641; Conservative 0;
```

[illegible]

QY	1311	gaagcagagagcccgagagatcgcgtcaactcagctcttgagatgacacgcgctctgc	1370
Db	24033	ggcgccgcccggccccaaggccggcgtacggcgccgcccgaaccgcgcaccttcgacggcgctggccga	24092
QY	1371	gcagcccgcgccgcctcgagaaagctcctcgacacgctg----cgcccgctgcgcccagga	1426
Db	24093	acggcccccggccgcccgggggggtrccgccttctccctggccttggaaacccctggggccggaggg	24152
QY	1427	tcgtacccgtgtgtgagacagagagcgaatcacaaactccgacacattccttgagacgcttca	1486
Db	24153	cacggccgcgcgcacggctggcgcgaagcggaaattctctcagccggcgccggcctggcctccctcga	24212
QY	1487	ccgagctctctgactactactaccacatglttcgattccctcctcgagagcgcgagctccgagc	1546
Db	24213	ccccgacacaggggggngc---ggacccttgccggccgacgttggaagcggcgacagccctgcg	24268
QY	1547	gcggcccatccgaagcttcattcctcgcgtgcgtctgtctctcgtccgcccgcgcacgagac	1606
Db	24269	gtgcggctggccgacgtccagatgcagatggaaacccggcttcgcgcctctacacactgggggtgcctccg	24328
QY	1607	aggtcatgtccgaggtgtcactccgcgcgcag	1638
Db	24329	ccgtactcttgcagacacatcccgacgtggac	24360

```

RESULT      8
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

[illegible]

[illegible]

Db 3928157 ACACCGGCGCGGAATCCGGGGTTTCCCGCGCCCGCGCGCGGAGGACACCGGCGCG 3928098

QY 1613 tgtccgaggtgtactctcgccgagacatcttgacagtgtgctctcgagagggcggaagc 1672

Db 3928097 AATCCGGCGTTTGGCCGCGCGTCCCGCGCATCGCGCTTGGCGCGCTACCGCGCTCGGT 3928038

QY 1673 gcaacagcgcgcagagacgcgtctggtgcagctgscgaaacgcgctgtgscagacgcgcggttcg 1732

Db 3928037 CCGCGGATCAAGGTGCGCGGGGGCCCGCTCACACACGGCGCCCGCGCGGTGCGCGGGCCCC 3927978

QY 1733 agaccgtccacccttggtctccaatgctctacaagacgagcagcagcgtctgtgcgtctctcg 1792

Db 3927977 CCGGCGCGCTTGATGCGCGCGCGCGCGCGCGCGCTTGGCGCGCGCGCGCGCGCTTGG 3927918

QY 1793 ccgagcgcgagcgcgtctacaagtgtgagggagag 1824

Db 3927917 CCGCGCTCAAGCGCGAGAGCGCCACGGCGCG 3927866

```

RESULT      9
US-09-130-114-2
: Sequence 2, Application US/09130114
: Patent No. 597607
: GENERAL INFORMATION:
: APPLICANT: Horlick, Robert A.
: APPLICANT: Damaq, Bassam B.
: APPLICANT: Robbins, Alan K.
: TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
: FILE REFERENCE: 0867/1D903U51
: CURRENT FILING DATE: 1998-08-06
: NUMBER OF SEQ. ID NOS.: 36
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 1931
: TYPE: DNA
: ORGANISM: EBNA
US-09-130-114-2

```

Query Match	Similarity	4.3%;	Score 91.8;	DB 2;	Length 1931;
Best Local Match	Similarity	45.9%;	Pred. No. 1.9e-07;		
Matches 351;	Conservative	0;	Mismatches 412;	Indels	2; Gaps
					1;
QY	806	tacccttgctgacgcgctccacaggcgagcgagatgcgaagatgcgcgcactactgcgacg	865		
DB	321	tcccctctctctgctctctctccgcgctctccgcgctctctctctctccgctctccgctctcc	380		
QY	866	agccctctgcgcgcgcgcgcgctcttcgcgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	925		
DB	381	tctctccgcgctctctgctctctctctccgcgctctctccgcgctctctccgcgctctctcc	440		
QY	926	cgcgccttcgc	985		
DB	441	cgcgtctctgctctctctctccgcgctctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	500		
QY	986	cgcacttcaacgcgcacacagagcgcactctgagagcgcttcgcgcgcgcgcgcgcgcgcgc	1045		
DB	501	tcccgcgtctctgctctctctccgcgctctctctctctctctctccgcgcgcgcgcgcgcgc	560		
QY	1046	tgcgtgcacttcgcgcacacagagggatgcagatgccccgcgcacttcctcagcgccctgcgc	1105		
DB	561	tctctccgcgctctctccgcgctctctccgcgctctctctctctctctccgcgctctctctcc	620		
QY	1106	tccgctccgc	1165		
DB	621	tctctccgcgctctctccgcgctctctctctctctccgcgcgcgcgcgcgcgcgcgcgcgcgc	680		
QY	1166	agacgc	1225		
DB	681	tctctccgcgctctctctccgc	740		

Qy	1226	tcgattccagatgacgagcgctctgtctgcgcgcacgctcgcgagctctgagcccttgatgac	1285
Db	741	tctctcccgctctctcccgctctctctgtctctcccgctctctcccgctctctctcccgctcc	800
Qy	1286	tgcagcgcgagggcgagaggaagccgaagagagcgccgaggttaatcgccgttaactcag	1345
Db	801	tccccgctctctctctctctctcccgctctcccgctctctctctctctctcccgctctcccg	860
Qy	1346	tcttcagatgagacgcgcgtctgcgcgcagccgcgcgcctctgagaaagctctcttgagacgc	1405
Db	861	tctctgcctctcccgctctctcccgctctctctgtctctctcccgctctcccgctctctctctcc	920
Qy	1406	tgcgcgcgtctgcgcgcgcacagatctgataccgctgctgagcagagagcgatatcaactcgc	1465
Db	921	ccgctctcccgctctctctgtctctctctcccgctctctctgc - tcccgctctctgtctcca	978
Qy	1466	gcacattctctgagacgcgtctcaaccgagctctctgcaactactcaacacgttcgattccc	1525
Db	979	ctctcgcgcacacgctctctctgtaactctcgcgcgcacagctctctcactcaactcgcgccca	1038
Qy	1526	tgcgagggcgagctcgcgcgcgcgcgcgcacatctcgaaagctcacaagc	1570
Db	1039	gtctctctcactactcgcgcgcgcgcgcacatctctgcaattcttcgc	1083

```

RESULT 10
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438U51
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

```

Query Match	4.28;	Score 89.2;	DB 4;	Length 15872;
Best Local Similarity	43.28;	Pred. No. 5.7e-07;		
Matches 580;	Conservative 0;	Mismatches 753;	Indels 9;	Gaps 3;

[illegible][illegible]

RESULT 11  
US-09-105-537-32

Sequence 32, Application US/09105537A  
Patent No. 6265202  
GENERAL INFORMATION:  
APPLICANT: Sherman, D. H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600.438US1  
CURRENT APPLICATION NUMBER: US/09/105.537A  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 32  
LENGTH: 11220  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-105-537-32

Query Match 4.1%; Score 87.2; DB 4; Length 11220;  
Best Local Similarity 42.7%; Pred. NO. 1.2e-06;  
Matches: 680; Conservative 0; Mismatches 896; Indels 14; Gaps 4;

QY 151 gcgagcgtcgggtacaaagtgccgcctccgcacatgagcgagactgagcaagaactgag 210  
DB 7819 gtggccctggcggcgagaggttcggcggtgacgagggcgcggttcggcttaccggcg 7878  
QY 211 cagctcagatgagtcagatgggagtgagcggtgagcgccgagccgcccagacagc 270  
DB 7879 cgtgcacgacgcccgtgacgagcggtgagcgcggtgagcggttcggcgcgagcg 7938  
QY 271 ttcgcac 330  
DB 7939 gctgcgcgttcgctcgtgagagcgatctcgtacgctgctgagcgacacacacacac 7998  
QY 331 gtcgaagacatgctgctgagagctcaacgacgacgacgacgacgacgacgacgac 390  
DB 7999 cggctcgcccgcccgcccgcccgcccgcccgcccgcccgcccgcccgcccgccg 8058  
QY 391 ctcaacgctccac 450  
DB 8059 ccggtgttcctccgacgacacacacacacacacacacacacacacacacacacac 8118  
QY 451 tccgctcctccac 510  
DB 8119 ttcacgac 8172  
QY 511 acgagcgcgacgac 569  
DB 8173 ggtgcgcgacgac 8232  
QY 570 cggagcgcgac 629  
DB 8233 gcgagcgcgacgac 8292  
QY 630 ctctctgtgagagctgc 689  
DB 8293 ----gagcgcgtgc 8348  
QY 690 gttcgtctgtgac 749  
DB 8349 cgc 8408  
QY 750 cggc 809  
DB 8409 gatgcagcgcgttcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 8468  
QY 810 ctgtgtgc 869  
DB 8469 cgaagcgcgttcgc 8528  
QY 870 cctgc 929

DB 8529 gggcctcgccgcttcgc 8588  
QY 930 ctgcgcac 989  
DB 8589 cggggaagc 8645  
QY 990 ctccaccccaac 1049  
DB 8646 cgcacacgc 8705  
QY 1050 cgaactcgac 1109  
DB 8706 cggctcgagc 8765  
QY 1110 tcccgagc 1169  
DB 8766 gctgagc 8825  
QY 1170 cgaagc 1229  
DB 8826 gctgctgc 8885  
QY 1230 ctccac 1289  
DB 8886 gagctcgc 8945  
QY 1290 gctgagagc 1349  
DB 8946 acagatccgc 9005  
QY 1350 cgaagatccgc 1409  
DB 9006 cggatgtac 9065  
QY 1410 cgc 1469  
DB 9066 cgc 9125  
QY 1470 attcctgc 1529  
DB 9126 gtaacgc 9185  
QY 1530 gggc 1589  
DB 9186 gttcgc 9245  
QY 1590 cgc 1649  
DB 9246 cctgagc 9305  
QY 1650 gttgc 1709  
DB 9306 catgagc 9365  
QY 1710 ccgagctggc 1741  
DB 9366 cgggagctggc 9397

RESULT 12  
US-09-105-537-5  
Sequence 5, Application US/09105537A  
Patent No. 6265202  
GENERAL INFORMATION:  
APPLICANT: Sherman, D. H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600.438US1  
CURRENT APPLICATION NUMBER: US/09/105.537A  
CURRENT FILING DATE: 1998-06-26









Db 15402 CAGAGGCGCGTGGAGCTCGGCCCTTTCGAGCTGCTCGCCGAGCGCTCCCGGCGTCCGAC 15461  
QY 268 agcttcgcccaccacccctcgccacgagaccgctgacactaaccccagacctgctcgt 327  
Db 15462 CAGATCCGCGCCGGAAGTGGGGCTCCGGCACCGGCTTTAAAGAGACTTCTTGAACGCCCTTG 15521  
QY 328 tgggtcgaagagctgctgctgagagctcaacgagcgccgcccgcctcccgccgcccgc 387  
Db 15522 GTCGGGCTCGGCGCTGCTGGAGGAGGAGCGGGGGCTTACCGGGGCTCCGAGCTCGGCGCG 15581  
QY 388 cagctcaacgctccacccctccaccgctccacgctcagggcagcgagctgactatcccg 447  
Db 15582 GAGTTCCTCTCCCGGCCCGCCAGTACTCTGGCGGCGACCGCCCGCCAGACGCTCGGCTG 15641  
QY 448 ccttcgctgacttcctccagcagcatctacgctcgtgagcgatcccccctcccgccgcg 507  
Db 15642 CACTTACCACTGCTGGGCGGACGCTGACCCGAGCGGCTGCGGAGCGCCAAAGTCCGCC 15701  
QY 508 ggcagcgcgccgagccgacccctgcccagctccgctcgagcga- tcccagcgagatgcgac 566  
Db 15702 GTGGCGCGGCGGAGGCGGCTGGGCTTACCCCAAGCACTAGAGGACCTGGACCGCGCGCG 15761  
QY 567 tggcgaggagcagacccctgctgctacatcctcctcctcgtcgtcgtggtggcgagc 626  
Db 15762 CAGATCATGCTGACATGAGCGCCCAACAGGTTTCACGCGCCGACGAGTTGGCGCGCGCG 15821  
QY 627 gaggctctgt 686  
Db 15822 ATCGACTGGAGCGCGGACTACCTCTCTGTGAGAGTGGCGCGCGCGCGCGCGCGCGCGCG 15881  
QY 687 ggc--gtctcgt 741  
Db 15882 TCCCGGATGCTACCGGCCACCGCGGAGCTGGCGGGGGGTCTTCGACCTTCCGCGCGCTG 15941  
QY 742 ctgagctgctgagagcgctgagcagagagaaactctccgagcgagcgctgtgtgtgtgtgt 801  
Db 15942 CCGCGGCTCTTGTAGAGAGAGCTGGTGGCTTGGCGGAGAACCGCGGAGCGGAGCTTCCAC 16001  
QY 802 cagatacccttctgt 861  
Db 16002 GCGCGTACTTCTTCCGACCGACCTGCGGAGGCGGAGCTGGTGTGCTTCCGCTCAGCGCC 16061  
QY 862 ggcgagggccctcgccgagctctctccgctccgcccagccg--acagctccctc 918  
Db 16062 TCCCGGACTGGGCGGCTCGGGGAGCCGCAAGAGCTGTGCGCGCGCCCAAGGCGGTG 16121  
QY 919 ctgagagcgccctcgagcactcctccacgagcacttctacgagctcgtgcccctactc 978  
Db 16122 CCGCGGCGGCGCTGT 16181  
QY 979 aagttcgggaacttccgacccgcaacagagccatccctggaagcgttcgcccgtcgccgcg 1038  
Db 16182 GAGGTCTCTGTCTACCGGATCAACACCACTATGATCCGGAGCGGCGGCGCTTACTCG 16241  
QY 1039 gtcagcgtctgacttcgactcaacagcagggatgagtgagcccgacttctccagcc 1098  
Db 16242 CTGCAAGGAGGCGCGGCTTACTGAGGAGGCGGCGGTTCACCGTCCAGCGGATCGCGCGC 16301  
QY 1099 ctgcgcctcgtcccgagcgccctccctcgttccgctca--ccggcgtcgcccccg 1156  
Db 16302 TCCGACACCATCAACCGCGGACCACTTCCGATCGGCTCAAGTCGGTCTGAAGGAAAGG 16361  
QY 1157 agcggagagagagcagcagcccgagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1213  
Db 16362 AGTTGACATGACCGGACACACCGCGGATCAAGCGCGGACCGCGCGCTTACGACG 16421  
QY 1214 aacacatccgctgactcagctcaacgagcgtcgtcgcgcaacgctcgaggaactgtg 1273  
Db 16422 TCGTCACTATCGGAGCGGCTGTGCGGACCATGCTCGGCTCGCAAGCAGG 16481  
QY 1274 agcgttcatgtcgagcgcgagggcgaggaaccggaacgagagcccgaggtatcg 1333

Db 16482 GCTTCCGATCATGCTGTGAGCGGTGCCACACCCCGCTTCCGCGCTGCGGAGTCCA 16541  
QY 1334 ccgtcaactcagctctcgagatgcaaccggtgctcgcgagcccgcgccctggaagaag 1393  
Db 16542 CCATCGGAGAGAGCGGT 16601  
QY 1394 tccctggcaccgtgagcgccgtgagcgccagagatcgtaacgtgtgtgtgtgtgtgtgtgtgtgt 1453  
Db 16602 TCGCCAACTGGCGAGCTTCCAGGAGCTCTCGCCAACTGACAGATTTCGACGCGGACGA 16661  
QY 1454 atcaaacctccgacatctcctggaacgcttcaacgagctctgcaactactatccacga 1513  
Db 16662 AGAGCAACTTGGGCTTCAATGTTCCACCGGGAGCGGAGGAGCGGAGCGGAGAGACCA 16721

Search completed: January 23, 2002, 05:43:16  
Job time: 23626 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:25:59 ; Search time 40.57 Seconds

(without alignments)  
23.736 Million cell updates/sec

Title: US-09-485-529-103

Perfect score: 76

Sequence: 1 LNA PPPPLPPAPQ 13

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

A\_Geneseq\_1101.\*  
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76	100.0	623	20	AAV02540
2	59	77.6	256	20	AAV02538
3	58	76.3	33	21	AA627119
4	57	75.0	182	22	AAAB80893
5	57	75.0	216	22	AAAB80892
6	57	75.0	223	18	AAW21733
7	57	75.0	282	22	AAAB95839
8	57	75.0	282	22	AAAB80896
9	57	75.0	321	22	AAAB95274
10	56	73.7	123	20	AAV02543
11	56	73.7	630	20	AAV02541

12	55	72.4	293	21	AAAB43674	Human Cancer assoc
13	55	72.4	313	22	AAAB79117	Corynebacterium gl
14	55	72.4	313	22	AAAB79242	Corynebacterium gl
15	55	72.4	347	21	AAAG14025	Arabidopsis thalia
16	55	72.4	356	21	AAAG14024	Arabidopsis thalia
17	55	72.4	359	21	AAAG14023	Arabidopsis thalia
18	55	72.4	438	22	AAAG90541	C glutamicum prote
19	55	72.4	446	12	AAAB1837	H2RIIRP nuclear ho
20	55	72.4	446	16	AAAB72484	H-2RIIRP. Not spe
21	55	72.4	448	14	AAAB39467	mxr-beta. Mus mu
22	54	71.1	90	21	AAAG19012	zee mays protein f
23	54	71.1	113	21	AAAG11750	Arabidopsis thalia
24	54	71.1	138	21	AAAG11749	Arabidopsis thalia
25	54	71.1	148	21	AAAB42921	Human ORF2585
26	54	71.1	167	21	AAAG11748	Arabidopsis thalia
27	54	71.1	168	18	AAAB26545	Mouse Huntington's
28	54	71.1	527	20	AAV29515	Human lung tumour
29	54	71.1	527	21	AAAB44492	Human lung tumour
30	54	71.1	580	22	AAAB60493	Human cell cycle a
31	54	71.1	2343	21	AAAB12453	Human HNRCP protei
32	54	71.1	2453	21	AAAB12454	Human HNRCP protei
33	53	69.7	55	20	AAV04937	Myobacterium spec
34	53	69.7	104	22	AAAM15554	Peptide #1988 enco
35	53	69.7	104	22	AAAM28043	Peptide #2080 enco
36	53	69.7	104	22	AAAM03303	Peptide #1985 enco
37	53	69.7	212	21	AAAG36695	Arabidopsis thalia
38	53	69.7	237	21	AAAG36694	Arabidopsis thalia
39	53	69.7	257	20	AAV25630	Phleum sp. allerger
40	53	69.7	270	21	AAAG36693	Arabidopsis thalia
41	53	69.7	285	20	AAV25633	Phleum sp. allerger
42	53	69.7	373	20	AAV04940	Myobacterium spec
43	53	69.7	390	21	AAAB58287	Lung cancer associ
44	53	69.7	427	20	AAV04941	Myobacterium spec
45	53	69.7	561	19	AAAB63701	Human hsk1 protein

#### ALIGNMENTS

RESULT 1	AAV02540	standard; Protein; 623 AA.
ID	AAV02540	
XX	AAV02540;	
AC	16-JUL-1999	(first entry)
DT	Protein encoded by wheat Rht clone 5a1 genomic sequence.	
XX		
XX	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;	
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;	
KW	paclobutrazol.	
XX		
OS	Triticum aestivum.	
XX		
PN	WO9909174-A1.	
XX	25-FEB-1999.	
PD		
XX	07-AUG-1998;	98WO-GB02383.
PF		
XX	13-AUG-1997;	97GB-0017192.
PR		
XX	(PLAN-) PLANT BIOSCIENCE LTD.	
PA	Harberd NP, Peng J, Richards DE;	
XX		
PI	WPI: 1999-181040/15.	
XX	N-PSDB; AAX36279.	
DR		
XX	New Triticum aestivum polynucleotides - encode a polypeptide which	
PT	provides inhibition of the growth of plants, which inhibition is	
PT	antagonised by gibberellin, used to confer a dwarf phenotype	

**PS** Disclosure; Fig 8b; 88pp; English.

**CC** The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in *Triticum aestivum*, inhibit growth of the plant. This growth inhibition is antagonised by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologue gene in the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol, e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds dwarf but let crop plants grow tall. The present sequence is encoded by the wheat Rht clone 5a1 genomic sequence.

**XX**

**SO** Sequence 623 AA;

**QY** 1 LNAPPPPLPAPAQ 13  
| | | | | | | | | |  
**Db** 109 lnapppplpapq 121

**RESULT 2**

**ID** AAY02538 standard; Protein; 256 AA.

**XX**

**AC** AAY02538;

**DT** 16-JUL-1999 (first entry)

**DE** Protein encoded by rice EST D39460 sequence.

**KW** Rht gene; homologue: *Triticum aestivum*; wheat; growth inhibition; antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis; paclobutrazol; rice; expressed sequence tag; EST.

**OS** *Oryza sativa*.

**XX**

**PN** MO9909174-A1.

**XX**

**PD** 25-FEB-1999.

**PF** 07-AUG-1998; 98WO-GB02383.

**PR** 13-AUG-1997; 97GB-0017192.

**PA** (PLAN-) PLANT BIOSCIENCE LTD.

**PI** Harberd NP, Peng J, Richards DE;

**DR** WPI: 1999-181040/15.  
N-PSTDB: AAX36277.

**PT** New *Triticum aestivum* polynucleotides - encode a polypeptide which provides inhibition of the growth of plants, which inhibition is antagonised by gibberellin, used to confer a dwarf phenotype

**PS** Claim 12; Fig 6b; 88pp; English.

**CC** The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in *Triticum aestivum*, inhibit growth of the plant. This growth inhibition is antagonised by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be

CC	used to produce Rht mutant plants which are dwarfed compared with wild-type; the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol, e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds dwarf but let crop plants grow tall. The present sequence is encoded by rice expressed sequence tag (EST) M6D39460, which is homologous to the wheat Rht gene.
CC	
XX	
SQ	Sequence 256 AA;
QY	1 LNAPPPPLPAP 12      ::
DB	109 lnappipppap 120
RESULT 3	
ID	AAG27119 standard; Protein: 33 AA.
XX	
AC	AAG27119;
XX	
DT	17-OCT-2000 (first entry)
XX	
DE	Zea mays protein fragment SEQ ID NO: 31833.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence; corn.
KW	
XX	
OS	Zea mays subsp. mays.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
FE	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.



PR 20-MAY-1999: 99US-0135124.  
PR 21-MAY-1999: 99US-0135553.  
PR 24-MAY-1999: 99US-01335629.  
PR 25-MAY-1999: 99US-0136621.  
PR 27-MAY-1999: 99US-0136392.  
PR 28-MAY-1999: 99US-0136782.  
PR 01-JUN-1999: 99US-0137222.  
PR 03-JUN-1999: 99US-0137528.  
PR 04-JUN-1999: 99US-0137502.  
PR 07-JUN-1999: 99US-0137724.  
PR 08-JUN-1999: 99US-0138094.  
PR 10-JUN-1999: 99US-0138540.  
PR 10-JUN-1999: 99US-0138847.  
PR 14-JUN-1999: 99US-0139119.  
PR 16-JUN-1999: 99US-0139452.  
PR 16-JUN-1999: 99US-0139453.  
PR 17-JUN-1999: 99US-0139492.  
PR 18-JUN-1999: 99US-0139454.  
PR 18-JUN-1999: 99US-0139455.  
PR 18-JUN-1999: 99US-0139456.  
PR 18-JUN-1999: 99US-0139457.  
PR 18-JUN-1999: 99US-0139458.  
PR 18-JUN-1999: 99US-0139459.  
PR 18-JUN-1999: 99US-0139460.  
PR 18-JUN-1999: 99US-0139461.  
PR 18-JUN-1999: 99US-0139462.  
PR 18-JUN-1999: 99US-0139463.  
PR 18-JUN-1999: 99US-0139750.  
PR 18-JUN-1999: 99US-0139763.  
PR 21-JUN-1999: 99US-0139817.  
PR 22-JUN-1999: 99US-0139899.  
PR 23-JUN-1999: 99US-0140353.  
PR 23-JUN-1999: 99US-0140354.  
PR 24-JUN-1999: 99US-0140695.  
PR 28-JUN-1999: 99US-0140823.  
PR 29-JUN-1999: 99US-0140991.  
PR 30-JUN-1999: 99US-0141287.  
PR 01-JUL-1999: 99US-0141842.  
PR 01-JUL-1999: 99US-0142154.  
PR 02-JUL-1999: 99US-0142055.  
PR 06-JUL-1999: 99US-0142390.  
PR 08-JUL-1999: 99US-0142803.  
PR 09-JUL-1999: 99US-0142920.  
PR 12-JUL-1999: 99US-0142977.  
PR 13-JUL-1999: 99US-0143542.  
PR 14-JUL-1999: 99US-0143624.  
PR 15-JUL-1999: 99US-0144005.  
PR 16-JUL-1999: 99US-0144085.  
PR 16-JUL-1999: 99US-0144086.  
PR 19-JUL-1999: 99US-0144325.  
PR 19-JUL-1999: 99US-0144331.  
PR 19-JUL-1999: 99US-0144332.  
PR 19-JUL-1999: 99US-0144333.  
PR 19-JUL-1999: 99US-0144334.  
PR 19-JUL-1999: 99US-0144335.  
PR 20-JUL-1999: 99US-0144352.  
PR 20-JUL-1999: 99US-0144632.  
PR 20-JUL-1999: 99US-0144814.  
PR 21-JUL-1999: 99US-0145086.  
PR 21-JUL-1999: 99US-0145088.  
PR 21-JUL-1999: 99US-0145088.  
PR 22-JUL-1999: 99US-0145085.  
PR 22-JUL-1999: 99US-0145087.  
PR 22-JUL-1999: 99US-0145089.  
PR 22-JUL-1999: 99US-0145192.  
PR 23-JUL-1999: 99US-0145145.  
PR 23-JUL-1999: 99US-0145218.  
PR 23-JUL-1999: 99US-0145224.  
PR 26-JUL-1999: 99US-0145276.  
PR 27-JUL-1999: 99US-0145913.  
PR 27-JUL-1999: 99US-0145918.  
PR 27-JUL-1999: 99US-0145919.  
PR 28-JUL-1999: 99US-0145951.  
  
PR 02-AUG-1999: 99US-0146386.  
PR 02-AUG-1999: 99US-0146388.  
PR 02-AUG-1999: 99US-0146389.  
PR 03-AUG-1999: 99US-0147038.  
PR 04-AUG-1999: 99US-0147204.  
PR 04-AUG-1999: 99US-0147302.  
PR 05-AUG-1999: 99US-0147192.  
PR 05-AUG-1999: 99US-0147260.  
PR 06-AUG-1999: 99US-0147303.  
PR 06-AUG-1999: 99US-0147416.  
PR 09-AUG-1999: 99US-0147493.  
PR 09-AUG-1999: 99US-0147935.  
PR 10-AUG-1999: 99US-0148171.  
PR 11-AUG-1999: 99US-0148319.  
PR 12-AUG-1999: 99US-0148341.  
PR 13-AUG-1999: 99US-0148565.  
PR 13-AUG-1999: 99US-0148684.  
PR 16-AUG-1999: 99US-0149368.  
PR 17-AUG-1999: 99US-0149175.  
PR 18-AUG-1999: 99US-0149426.  
PR 20-AUG-1999: 99US-0149722.  
PR 20-AUG-1999: 99US-0149723.  
PR 20-AUG-1999: 99US-0149929.  
PR 23-AUG-1999: 99US-0149902.  
PR 23-AUG-1999: 99US-0149930.  
PR 25-AUG-1999: 99US-0150566.  
PR 26-AUG-1999: 99US-0150884.  
PR 27-AUG-1999: 99US-0151065.  
PR 27-AUG-1999: 99US-0151066.  
PR 27-AUG-1999: 99US-0151080.  
PR 30-AUG-1999: 99US-0151303.  
PR 31-AUG-1999: 99US-0151438.  
PR 01-SEP-1999: 99US-0151930.  
PR 07-SEP-1999: 99US-0152363.  
PR 10-SEP-1999: 99US-0153070.  
PR 13-SEP-1999: 99US-0153758.  
PR 15-SEP-1999: 99US-0154018.  
PR 16-SEP-1999: 99US-0154039.  
PR 20-SEP-1999: 99US-0154779.  
PR 22-SEP-1999: 99US-0155139.  
PR 23-SEP-1999: 99US-0155486.  
PR 24-SEP-1999: 99US-0155659.  
PR 28-SEP-1999: 99US-0156458.  
PR 29-SEP-1999: 99US-0156596.  
PR 04-OCT-1999: 99US-0157117.  
PR 05-OCT-1999: 99US-0157753.  
PR 06-OCT-1999: 99US-0157865.  
PR 07-OCT-1999: 99US-0158029.  
PR 08-OCT-1999: 99US-0158232.  
PR 12-OCT-1999: 99US-0158369.  
PR 13-OCT-1999: 99US-0159293.  
PR 13-OCT-1999: 99US-0159294.  
PR 13-OCT-1999: 99US-0159295.  
PR 13-OCT-1999: 99US-0159329.  
PR 14-OCT-1999: 99US-0159330.  
PR 14-OCT-1999: 99US-0159331.  
PR 14-OCT-1999: 99US-0159637.  
PR 14-OCT-1999: 99US-0159638.  
PR 18-OCT-1999: 99US-0159584.  
PR 21-OCT-1999: 99US-0160741.  
PR 21-OCT-1999: 99US-0160767.  
PR 21-OCT-1999: 99US-0160768.  
PR 21-OCT-1999: 99US-0160770.  
PR 21-OCT-1999: 99US-0160814.  
PR 21-OCT-1999: 99US-0160815.  
PR 22-OCT-1999: 99US-0160980.  
PR 22-OCT-1999: 99US-0160981.  
PR 22-OCT-1999: 99US-0160989.  
PR 25-OCT-1999: 99US-0161404.  
PR 25-OCT-1999: 99US-0161405.  
PR 25-OCT-1999: 99US-0161406.  
PR 26-OCT-1999: 99US-0161359.  
PR 26-OCT-1999: 99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 76.3%; Score 58; DB 21; Length 33;

Best Local Similarity 90.0%; Pred. No. 2;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13  
13 PPPPLPPAPR 22

RESULT 4

AAAB0893 standard; protein; 182 AA.

AC AAB80893;

DT 30-MAY-2001 (first entry)

DE Human ATFx leucine zipper domain.

KW Human; ATFx; gamma aminobutyric acid B receptor; GABA B receptor;

transcription factor; leucine zipper.

OS Homo sapiens.

PN WO200116596-A2.

PD 08-MAR-2001.

PF 31-AUG-2000; 2000WO-GB03361.

PR 31-AUG-1999; 99GB-0020569.

PR 12-JAN-2000; 2000GB-0000516.

PA (GLAX ) GLAXO GROUP LTD.

PI White J, Wise A, Marshall F;

DR WPI: 2001-265904/27.

DR N-PSDB; AAF77878.

PT Identifying modulators of gamma aminobutyric acid(GABA)-B receptor  
PT mediated activity by monitoring the interaction between GABAB receptor  
PT and the CREB/ATF transcription factors in the presence of a test  
PT substance

PS Claim 14; Page 64-65; 71pp; English.

XX The present invention relates to a method for identifying a modulator of  
CC gamma aminobutyric acid<sub>B</sub> (GABA<sub>B</sub>) receptor-mediated activity, by  
CC monitoring the interaction between a CREB/ATF transcription factor  
CC capable of binding to GABA<sub>B</sub> receptor. The present sequence is the  
CC leucine zipper domain of human ATFx (see AAB80892), which was used in the  
CC method of the present invention. ATFx is a member of the CREB/ATF family  
CC of transcription factors. ATFx contains a bZIP domain, by which it  
CC interacts with the coiled coil domain of GABA<sub>B</sub> receptor. Modulators of  
CC GABA<sub>B</sub> receptor activity are useful for treating central nervous system  
CC or peripheral nervous system disorders.

XX Sequence 182 AA;

Query Match 75.0%; Score 57; DB 22; Length 182;

Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
1111111111

DB 31 PPPPLPPAP 39

RESULT 5

AAAB0892 standard; protein; 216 AA.

AC AAB80892;

DT 30-MAY-2001 (first entry)

DE Human ATFx.

KW Human; ATFx; gamma aminobutyric acid B receptor; GABA B receptor;

transcription factor.

OS Homo sapiens.

PN WO200116596-A2.

PD 08-MAR-2001.

PF 31-AUG-2000; 2000WO-GB03361.

PR 31-AUG-1999; 99GB-0020569.

PR 12-JAN-2000; 2000GB-0000516.

PA (GLAX ) GLAXO GROUP LTD.

PI White J, Wise A, Marshall F;

DR WPI: 2001-265904/27.

DR N-PSDB; AAF77877.

PT Identifying modulators of gamma aminobutyric acid(GABA)-B receptor  
PT mediated activity by monitoring the interaction between GABAB receptor  
PT and the CREB/ATF transcription factors in the presence of a test  
PT substance

PS Claim 14; Fig 1; 71pp; English.

XX The present invention relates to a method for identifying a modulator of  
CC gamma aminobutyric acid<sub>B</sub> (GABA<sub>B</sub>) receptor-mediated activity, by  
CC monitoring the interaction between a CREB/ATF transcription factor  
CC capable of binding to GABA<sub>B</sub> receptor. The present sequence is human  
CC ATFx, which was used in the method of the present invention. ATFx is a  
CC member of the CREB/ATF family of transcription factors. ATFx contains a  
CC bZIP domain, by which it interacts with the coiled coil domain of GABA<sub>B</sub>  
CC receptor. Modulators of GABA<sub>B</sub> receptor activity are useful for treating  
CC central nervous system or peripheral nervous system disorders.

XX Sequence 216 AA;

Query Match 75.0%; Score 57; DB 22; Length 216;

Best Local Similarity 100.0%; Pred. No. 13;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
1111111111

DB 65 PPPPLPPAP 73

RESULT 6

AAW21733 standard; Protein; 223 AA.

AC AAW21733;

DT 01-OCT-1997 (first entry)

DE NIP-1 encoded by clone 59.

XX

KM NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;  
 KM cell division; proliferation; antibody; Ab; detection;  
 KM malignant cell growth.  
 OS Homo sapiens.  
 XX  
 XX WO9640917-A1.  
 XX  
 PD 19-DEC-1996.  
 XX  
 XX 07-JUN-1996; 96WO-US09504.  
 XX  
 XX 07-JUN-1995; 95US-0478408.  
 XX  
 XX (UYVA ) UNIV YALE.  
 XX  
 XX McPherson SMG, Snyder MP;  
 XX  
 XX MPI: 1997-077270/07.  
 DR N-PSDB; AAT77784.  
 XX  
 XX New nucleic acid encoding nuclear mitotic appts. interacting  
 PT proteins - useful for modulating cell division and proliferation and  
 PR in diagnosis  
 XX  
 XX Claim 28; Page 57-58; 78pp; English.  
 PS  
 XX The sequences given in AAW21733-35 represent NuMA binding proteins (NuMA  
 CC - nuclear mitotic apparatus). These protein sequences were identified  
 CC using the fusion proteins given in AAW21731-32. Compounds which  
 CC interfere with the interaction of NuMA with a known NIP (NuMA interacting  
 CC protein) are used to modulate cell division and/or proliferation. Ab,  
 CC raised conventionally using NIP-1 or -2 as immunogen, are used to detect  
 CC NIP (or their complexes) and to block their activity for diagnostic or  
 CC therapeutic use, e.g. to detect defective NuMA or NIP which may be  
 CC markers for aberrant (including malignant) cell growth (which can also  
 CC be detected by nucleic acid sequencing). Also where malignancy is  
 CC related to defects in NuMA or NIP, it can be treated by administration  
 CC of the appropriate functional protein. This protein is rich in  
 CC proline residues.  
 CC  
 CC Sequence 223 AA:  
 SQ

Query Match 75.0%; Score 57; DB 18; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
 Db 2 PPPPLPPAP 10  
 |||||

RESULT 7  
 AAB95859  
 ID AAB95859 standard; Protein; 282 AA.  
 XX  
 AC AAB95859;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:18922.  
 XX  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 KM  
 XX Homo sapiens.  
 OS  
 XX EP1074617-A2.  
 PN  
 XX 07-FEB-2001.  
 PD  
 XX 28-JUL-2000; 2000EP-0116126.  
 PF  
 XX

PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 XX (HELI-) HELIX RES INST.  
 XX  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 PI  
 DR MPI: 2001-318749/34.  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 XX  
 PS Claim 8; SEQ ID 18922; 2537pp + CD ROM; English.  
 XX  
 XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides, and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 CC Sequence 282 AA:  
 SQ

Query Match 75.0%; Score 57; DB 22; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
 Db 131 PPPPLPPAP 139  
 |||||

RESULT 8  
 AAB80896  
 ID AAB80896 standard; Protein; 282 AA.  
 XX  
 AC AAB80896;  
 XX  
 DT 30-MAY-2001 (first entry)  
 XX  
 DE Human ATFX with N-terminal extension.  
 XX  
 XX Human; ATFX; gamma aminobutyric acid B receptor; GABA B receptor;  
 KM transcription factor.  
 KM  
 XX Homo sapiens.  
 OS  
 XX WO200116596-A2.  
 PN  
 XX 08-MAR-2001.  
 PD  
 XX

XX 31-AUG-2000; 2000WO-GB03361.  
 PF 31-AUG-1999; 99GB-0020569.  
 XX 12-JAN-2000; 2000GB-0000516.  
 PR (GLAXO ) GLAXO GROUP LTD.  
 XX  
 PA White J, Wise A, Marshall F;  
 XX  
 PI WPI: 2001-265904/27.  
 DR N-PSDB: AAF77880.  
 XX  
 PR Identifying modulators of gamma aminobutyric acid(GABA)-B receptor  
 PT mediated activity by monitoring the interaction between GABAB receptor  
 PR and the CREB/ATF transcription factors in the presence of a test  
 PR substance -  
 PS  
 XX Claim 14; Fig 9; 71pp; English.  
 PS  
 XX The present invention relates to a method for identifying a modulator of  
 CC gamma aminobutyric acid\_B (GABA\_B) receptor-mediated activity, by  
 CC monitoring the interaction between a CREB/ATF transcription factor  
 CC capable of binding to GABA\_B receptor. The present sequence is human  
 CC ATRX, which was used in the method of the present invention. ATRX is a  
 CC member of the CREB/ATF family of transcription factors. ATRX contains a  
 CC bZIP domain, by which it interacts with the coiled coil domain of GABA\_B  
 CC receptor. Modulators of GABA\_B receptor activity are useful for treating  
 CC central nervous system or peripheral nervous system disorders. The  
 CC present sequence has an N-terminal extension compared to AAB80892.  
 CC  
 XX Sequence 282 AA;  
 SQ

Query Match 75.0%; Score 57; DB 22; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 4 PPPPLPPAP 12  
 |||||  
 DB 131 PPPPLPPAP 139

RESULT 9  
 AAB95274  
 ID AAB95274 standard; Protein: 321 AA.  
 XX  
 AC AAB95274;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:17476.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX

XX WPI: 2001-318749/34.  
 DR  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 XX  
 PS Claim 8; SEQ ID 17476; 2537pp + CD ROM; English.  
 XX

CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 XX Sequence 321 AA;  
 SQ

Query Match 75.0%; Score 57; DB 22; Length 321;  
 Best Local Similarity 76.9%; Pred. No. 18;  
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 OY 1 LNAPPPPLPPAPQ 13  
 | |||| |  
 DB 121 LPppppppppapq 133

RESULT 10  
 AAY02543  
 ID AAY02543 standard; Protein: 123 AA.  
 XX  
 AC AAY02543;  
 XX  
 DT 16-JUL-1999 (first entry)  
 XX  
 DE Protein encoded by the partial sequence of the maize D8-2023 allele.  
 XX  
 KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
 KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
 KW paclobutrazol; maize.  
 XX  
 OS Zea mays.  
 XX  
 PN WO9909174-A1.  
 XX  
 PD 25-FEB-1999.  
 XX  
 PF 07-AUG-1998; 98WO-GB02383.  
 XX  
 PR 13-AUG-1997; 97GB-0017192.  
 XX  
 PA (PLAN-) PLANT BIOSCIENCE LTD.  
 XX  
 PI Harberd NP, Peng J, Richards DE;  
 XX

DR	WP1: 1999-181040/15.
DR	N-PSDB; AAX36282.
XX	
PT	New Triticum aestivum polynucleotides - encode a polypeptide which
PT	provides inhibition of the growth of plants, which inhibition is
PT	antagonised by gibberellin, used to confer a dwarf phenotype
XX	
PS	Disclosure; Fig 11d; 88pp; English.
XX	
CC	The specification describes polypeptides encoded by the Rht gene (and
CC	its homologues) that, when expressed in Triticum aestivum, inhibit
CC	growth of the plant. This growth inhibition is antagonised by
CC	gibberellin. The products can be used to provide Rht expression in
CC	plants, conferring a dwarf phenotype on a plant which is correctable
CC	by treatment with gibberellin. In addition, the products can be
CC	used to produce Rht mutant plants which are dwarfed compared with
CC	wild-type, the dwarfing being gibberellin-insensitive. Taller plants
CC	may be made by knocking out Rht or the relevant homologous gene in
CC	the plant of interest. Plants may be made which are resistant to
CC	compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
CC	e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
CC	dwarf but let crop plants grow tall. The present sequence is encoded by
CC	the partial sequence of the maize D8-2023 allele.
XX	
SQ	Sequence 123 AA;
	Query Match 73.7%; Score 56; DB 20; Length 123;
	Best Local Similarity 90.9%; Pred. No. 10;
	Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0.
OY	1 LNAPPPPLPAA 11 
Db	82 lnapppaplpaa 92
RESULT 11	
AY02541	
ID	AA02541 standard; Protein; 630 AA.
XX	
AC	AA02541;
XX	
DT	16-JUL-1999 (first entry)
XX	
DE	Protein encoded by maize 1a1 genomic clone sequence.
KW	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KM	paclobutrazol; maize.
XX	
OS	Zea mays.
XX	
PN	MO90909174-A1.
XX	
PD	25-FEB-1999.
XX	
PF	07-AUG-1998; 98WO-CB02383.
XX	
PR	13-AUG-1997; 97GB-0017192.
XX	
PA	(PLAN-) PLANT BIOSCIENCE LTD.
XX	
PI	Harberd NP, Peng J, Richards DE;
XX	
DR	WP1: 1999-181040/15.
DR	N-PSDB; AAX36280.
XX	
PT	New Triticum aestivum polynucleotides - encode a polypeptide which
PT	provides inhibition of the growth of plants, which inhibition is
PT	antagonised by gibberellin, used to confer a dwarf phenotype
XX	
BS	Disclosure; Fig 9b; 86pp; English.
XX	

CC	The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in Triticum Aestivum, inhibit growth of the plant. This growth inhibition is antagonised by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type. The dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in CC the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol, e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds dwarf but let crop plants grow tall. The present sequence is encoded by CC the maize 1al genomic clone sequence.
SQ	Sequence   630 AA;
OY	1 INAPPPPIPPA 11             Db         110 InappapiPpa 120
DE	RESULT_12 AAB43674 ID   AAB43674 standard; Protein: 293 AA. XX AC   AAB43674; XX DT   08-FEB-2001 (first entry) XX
DE	Human cancer associated protein sequence SEQ ID NO:1119. XX KW   Human; cancer associated gene; cancer antigen; detection; cancer diagnosis; cytostatic; Proliferative; vulnerability; immunomodulator; antiadipogenic; antiaesthetic; antirheumatic; antithrombotic; antiviral; antiflammatoy; antihypertoid; antiallergic; antibacterial; cardiac; dermatological; neuroprotective; thrombolytic; coagulant; noctropic; vasotropic; antipsoriatic; angiologenic; gene therapy; inflammation; immune disorder; haematopoietic cell disorder; autoimmune disorder; allergic reaction; graft versus host disease; organ rejection; haemostatic; thrombolytic; cardiovascular disorder; infection; neurological disease; drug screening. KW KM XX OS   Homo sapiens. OS PN   WO200055350-A1. PN PD   21-SEP-2000. PD PF   08-MAR-2000; 2000WO-USO5882. PF PR -  12-MAR-1999;      99US-O124270. PR PA   (HUMA-) HUMAN GENOME SCI INC. PA PI   Rosen CA, Ruben SM; PI DR   WPI: 2000-587533/75. DR N-PSDB; AAC77883. XX PT   Novel isolated nucleic acids comprising sequences encoding peptides useful for treating or diagnosing e.g. cancer - PT XX PS   Claim 11; Page 1730-1731; 2352pp; English. PS XX AC   AAC77607 to AAC78448 encode the human cancer associated proteins given in AAB43398 to AAB44239. The proteins can have activities based on the tissues and cells the genes are expressed in. Example of activities

CC include: cytostatic; proliferative; vulnerary; immunomodulator;  
 CC antidiabetic; antisthmatic; antirheumatic; antiallergic;  
 CC antiinflammatory; antihypertoid; antiallergic; antibacterial; antiviral;  
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
 CC neurotropic; vasotropic; antipruritic and antiangiogenic. The  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC rejection, modulate haemostatic or thrombolytic activity, modulate  
 CC inflammation, cancers, cardiovascular disorders, neurological disease and  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
 CC the present invention.  
 CC  
 XX Sequence 293 AA:

Query Match 72.4%; Score 55; DB 21; Length 293;  
 Best Local Similarity 90.0%; Pred. No. 28;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13  
 |||||  
 Db 16 PPPPPPPAPQ 25

RESULT 13  
 AAB79117  
 ID AAB79117 standard; Protein: 313 AA.

XX AAB79117;  
 DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum HA protein sequence SEQ ID NO:190.

XX Corynebacterium glutamicum; homeostasis; adaptation; HA protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;  
 KW diagnosis; Corynebacterium diptheriae; genetic engineering;  
 KW Brevibacterium; environmental condition.

OS Corynebacterium glutamicum.

PN WO200100842-A2.

XX 04-JAN-2001.

PF 23-JUN-2000; 2000WO-IB00911.

XX 25-JUN-1999; 99US-0141031.  
 PR 08-JUL-1999; 99DE-1031636.  
 PR 09-JUL-1999; 99DE-1032125.  
 PR 09-JUL-1999; 99DE-1032126.  
 PR 09-JUL-1999; 99DE-1032127.  
 PR 09-JUL-1999; 99DE-1032128.  
 PR 09-JUL-1999; 99DE-1032129.  
 PR 09-JUL-1999; 99DE-1032226.  
 PR 14-JUL-1999; 99DE-1032920.  
 PR 14-JUL-1999; 99DE-1032922.  
 PR 14-JUL-1999; 99DE-1032924.  
 PR 14-JUL-1999; 99DE-1032928.  
 PR 14-JUL-1999; 99DE-1032930.  
 PR 14-JUL-1999; 99DE-1032933.  
 PR 14-JUL-1999; 99DE-1032935.  
 PR 14-JUL-1999; 99DE-1032973.

PR 14-JUL-1999; 99DE-1033002.  
 PR 14-JUL-1999; 99DE-1033003.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 31-AUG-1999; 99DE-1041390.  
 PR 31-AUG-1999; 99DE-1041391.  
 PR 03-SEP-1999; 99DE-1042088.

PA (BADI ) BASF AG.

XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;

XX WPI; 2001-061974/07.

DR N-PSDB; AAF71232.

PT New isolated Corynebacterium glutamicum nucleic acid for production or  
 PT modulation of production of fine chemicals such as amino acids,  
 PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins  
 PT or enzymes -

PS Claim 20; Page 399-400; 712pp; English.

CC AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis  
 CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The  
 CC C. glutamicum HA genes (I) can be used in vectors for expression in host  
 CC cells and production of fine chemicals, such as, an organic acid,  
 CC proteinogenic or nonproteinogenic amino acid (preferred), purine or  
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated  
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,  
 CC polypeptide or enzyme. The amino acids produced can be lysine, glutamine,  
 CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,  
 CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,  
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can  
 CC be modulated. The presence of (I) or HA proteins encoded by them are  
 CC used for diagnosing the presence or activity of Corynebacterium  
 CC diptheriae. (I) can be used to map the C. glutamicum genome or can be  
 CC used as markers for genetically engineered Corynebacterium or  
 CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain  
 CC homeostasis in C. glutamicum or help the microorganism to adapt to  
 CC different environmental conditions.

SO Sequence 313 AA;

Query Match 72.4%; Score 55; DB 22; Length 313;  
 Best Local Similarity 90.0%; Pred. No. 30;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LNAPPPLPP 10  
 |:|||||  
 Db 288 lappplpp 297

RESULT 14  
 AAB79242  
 ID AAB79242 standard; Protein: 313 AA.

XX AAB79242;

DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum HA protein sequence SEQ ID NO:440.

XX Corynebacterium glutamicum; homeostasis; adaptation; HA protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;  
 KW diagnosis; Corynebacterium diptheriae; genetic engineering;  
 KW Brevibacterium; environmental condition.

OS Corynebacterium glutamicum.  
XX  
PN W0200100842-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 23-JUN-2000; 2000MO-IB00911.  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 08-JUL-1999; 99DE-1031636.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032126.  
PR 09-JUL-1999; 99DE-1032127.  
PR 09-JUL-1999; 99DE-1032128.  
PR 09-JUL-1999; 99DE-1032129.  
PR 09-JUL-1999; 99DE-1032226.  
PR 14-JUL-1999; 99DE-1032920.  
PR 14-JUL-1999; 99DE-1032922.  
PR 14-JUL-1999; 99DE-1032924.  
PR 14-JUL-1999; 99DE-1032928.  
PR 14-JUL-1999; 99DE-1032930.  
PR 14-JUL-1999; 99DE-1032933.  
PR 14-JUL-1999; 99DE-1032935.  
PR 14-JUL-1999; 99DE-1033002.  
PR 14-JUL-1999; 99DE-1033003.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041390.  
PR 31-AUG-1999; 99DE-1041391.  
PR 03-SEP-1999; 99DE-1042088.  
XX  
PA (BADI ) BASF AG.  
XX  
PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
XX  
DR WPI: 2001-061974/07.  
XX  
DR N-PSDB; AAF71357.  
XX  
PT New isolated Corynebacterium glutamicum nucleic acid for production or  
PT modulation of production of fine chemicals such as amino acids,  
PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins  
PT or enzymes -  
XX  
PS Claim 20; Page 711-712; 712pp; English.  
XX  
XX AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis  
CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The  
CC C. glutamicum HA genes (I) can be used in vectors for expression in host  
CC cells and production of fine chemicals, such as, an organic acid,  
CC proteinogenic or nonproteinogenic amino acid (preferred), putine or  
CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated  
CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,  
CC polyketide or enzyme. The amino acids produced can be lysine, glutamine,  
CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,  
CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,  
CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can  
CC be modulated. The presence of (I) or HA proteins encoded by them are  
CC used for diagnosing the presence or activity of Corynebacterium  
CC diptheriae. (I) can be used to map the C. glutamicum genome or can be  
CC used as markers for genetically engineered Corynebacterium or  
CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain  
CC homeostasis in C. glutamicum or help the microorganism to adapt to  
CC different environmental conditions.  
XX  
SQ Sequence 313 AA:

Query Match 72.4%; Score 55; DB 22; Length 313;  
Best Local Similarity 90.0%; Pred. No. 30;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LNAPPPPLPP 10  
Db 288 lapppplpp 297  
RESULT 15  
AGI4025  
ID AGI4025 standard; Protein; 347 AA.  
AC AGI4025;  
XX  
DE 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 13734.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128274.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132485.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139839.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140355.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140981.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161982.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 72.4%; Score 55; DB 21; Length 347;

Best Local Similarity 81.8%; Pred. NO. 32;

Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 NAPPPLPPAP 12  
| | | | |  
Db 151 nrppppppap 161



Search completed: January 22, 2002, 16:26:00  
Job time: 140 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:24:46 ; Search time 26.73 seconds  
(without alignments)  
37.047 Million cell updates/sec

Title: US-09-485-529-103

Perfect score: 76

Sequence: 1 LNAPPPLPPAPQ 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR:68:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	56	73.7	599	2	T10798	pherophorin-S - Vo
2	55	72.4	267	2	T15645	hypothetical prote
3	55	72.4	367	2	T39172	cyclin-dependent k
4	55	72.4	446	2	A34418	H-2 region II bind
5	55	72.4	448	2	D41727	retinoid X recepto
6	55	72.4	451	2	A41651	retinoic acid rece
7	55	72.4	520	2	T84718	RXR-beta1 isoform
8	54	71.1	138	2	C96734	hypothetical prote
9	54	71.1	998	2	T30930	hypothetical prote
10	54	71.1	2453	2	S60254	nuclear receptor c
11	54	71.1	3119	2	T49729	HD protein - mouse
12	53	69.7	257	2	S32101	PHL5A protein - c
13	53	69.7	373	2	A70856	proliferating cell
14	53	69.7	731	2	T19721	hypothetical prote
15	52	68.4	238	2	T32889	hypothetical prote
16	52	68.4	281	2	T38707	Fas ligand - human
17	52	68.4	359	2	T13478	hypothetical prote
18	52	68.4	389	2	S27200	proline-rich prote
19	52	68.4	497	2	T15812	hypothetical prote
20	52	68.4	543	2	S25128	61k protein - Auto
21	52	68.4	561	2	T03908	unknown protein P2
22	52	68.4	596	2	T03908	hypothetical prote
23	52	68.4	699	2	A96529	hypothetical prote
24	52	68.4	798	2	T19864	hypothetical prote
25	52	68.4	1132	2	A35098	MHC class III hist
26	52	68.4	1151	2	T18535	high molecular mas
27	52	68.4	1264	2	A36858	G2R protein - vari
28	52	68.4	1308	2	T15280	hypothetical prote
29	52	68.4	1388	2	T00063	hypothetical prote

30	52	68.4	1460	1	EDBEIF	immediate-early pr
31	52	68.4	1560	2	T00080	hypothetical prote
32	52	68.4	1733	1	B45344	probable nuclear a
33	52	68.4	1872	2	T24683	hypothetical prote
34	52	68.4	1896	2	B72175	D1SR protein - var
35	52	68.4	1897	2	T28621	hypothetical prote
36	52	68.4	1958	2	B40505	hypothetical prote
37	52	68.4	2123	2	T86348	hypothetical prote
38	51	67.1	281	2	T29150	hypothetical prote
39	51	67.1	294	2	A55477	Survival motor neu
40	51	67.1	399	2	A54754	homeobox protein -
41	51	67.1	433	2	T07910	hydroxyproline-ric
42	51	67.1	541	2	T19304	hypothetical prote
43	51	67.1	564	2	T45866	hypothetical prote
44	51	67.1	574	2	T05964	probable low-affin
45	51	67.1	666	2	B70803	hypothetical prote

#### ALIGNMENTS

RESULT 1  
T10798  
pherophorin-S - Volvox carteri  
C:Species: Volvox carteri  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: T10798  
R:Godl, R.; Hallmann, A.; Wenzl, S.; Sumper, M.  
EMBO J. 16, 25-34, 1997  
A:Title: Differential targeting of closely related ECM-glycoproteins: The pherophori  
A:Reference number: Z17154, PMID:97162277  
A:Accession: T10798  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-599 <GOD>  
A:Molecule type: DNA  
A:References: EMBL:Y07752; NID:q1655698; PIDN:CAA69032.1; PID:q1655699  
A:Experimental source: strain HK 10; sub-species Nagariensis  
A>Note: In contrast to the other pherophorins, pherophorin-S is targeted to the cell-  
C:Keywords: extracellular matrix; glycoprotein; pheromone

Query Match 73.7%; Score 56; DB 2; Length 599;  
Best Local Similarity 81.8%; Pred. No. 16;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 NAPPPLPPAP 12  
Db 216 NAPPPLPPAP 226

RESULT 2  
T15645  
hypothetical protein C26F1.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T15645  
R:Geisler, C.; Steliyes, L.; Bradshaw, H.  
submitted to the EMBL Data Library, March 1996  
A:Description: The sequence of C. elegans cosmid C26F1.  
A:Reference number: Z18381  
A:Accession: T15645  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-267 <GEIS>  
A:Cross-references: EMBL:U53148; NID:q1255375; PID:q1255384; PIDN:AA37073.1; GSPDB:G  
A:Experimental source: strain Bristol N2; clone C26F1  
C:Genetics:  
A:Gene: C26F1.1  
A:Map position: 5  
A:Introns: 54/3; 69/2; 85/2; 156/2; 171/3; 192/3

Query Match 72.4%; Score 55; DB 2; Length 267;

Best Local Similarity 90.0%; Pred. No. 9.5;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4 PPPPLPPAPQ 13  
|||||  
Db 90 PPPPPPPAPQ 99

## RESULT 3

139172  
cyclin-dependent kinase 5 activator isoform p391 - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Mar-1996 #sequence\_revision 01-Mar-1996 #text\_change 21-Jul-2000  
C:Accession: J39172  
R:Tang, D.; Yeung, J.; Lee, K.; Matsushita, M.; Matsui, H.; Tomizawa, K.; Hatake, O.; Waj  
J. Biol. Chem. 270, 26897-26903, 1995  
A:Title: An isoform of the neuronal cyclin-dependent kinase 5 (cdk5) activator.  
A:Reference number: 139172; MUID:96070784  
A:Accession: 139172  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-367 <RES>  
A:Cross-references: EMBL:U34051; NID:g1063622; PIDN:AAC50278.1; PID:g1063623  
C:Superfamily: tau-protein kinase II 23K chain

Query Match 72.4%; Score 55; DB 2; Length 367;  
Best Local Similarity 90.0%; Pred. No. 13;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAPQ 13  
|||||  
Db 155 PPPPPPPAPQ 164

## RESULT 4

A34418  
H-2 region II binding protein - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Jun-1990 #sequence\_revision 09-Oct-1992 #text\_change 20-Sep-1999  
C:Accession: A34418  
R:Hamada, K.; Gleason, S.L.; Levy, B.Z.; Hirschfeld, S.; Appella, E.; Ozato, K.  
Proc. Natl. Acad. Sci. U.S.A. 86, 8289-8293, 1989  
A:Title: H-2RIIBP, a member of the nuclear hormone receptor superfamily that binds to b  
A:Reference number: A34418; MUID:90046782  
A:Accession: A34418  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-446 <HAM>  
A:Cross-references: GB:M26804; NID:g193740; PIDN:AAA37772.1; PID:g387184  
C:Superfamily: unassigned erba-related proteins; erba transforming protein homology  
C:Keywords: DNA binding; transcription regulation; zinc finger  
F:116-367/Domain: erba transforming protein homology <ERBA>

Query Match 72.4%; Score 55; DB 2; Length 446;  
Best Local Similarity 72.7%; Pred. No. 16;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 NAPPPLPPAP 12  
:|||||  
Db 31 SAPPPMPPPP 41

## RESULT 5

D41727  
retinoid X receptor beta - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 17-Mar-1999  
C:Accession: D41727  
R:Leid, M.; Kastner, P.; Lyons, R.; Nakshtri, H.; Saunders, M.; Zacharski, T.; Chen,  
Cell 68, 377-395, 1992  
A:Title: Purification, cloning, and RXR identity of the HeLa cell factor with which RXR

A:Reference number: A41727; MUID:92127595  
A:Accession: D41727  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-448 <LEI>  
A:Experimental source: liver  
A>Note: sequence extracted from NCBI backbone (NCBI:77547)  
C:Genetics:  
A:Start codon: CTG  
C:Superfamily: unassigned erba-related proteins; erba transforming protein homology  
C:Keywords: zinc finger  
F:118-364/Domain: erba transforming protein homology <ERBA>

Query Match 72.4%; Score 55; DB 2; Length 448;  
Best Local Similarity 72.7%; Pred. No. 16;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 NAPPPLPPAP 12  
:|||||  
Db 33 SAPPPMPPPP 43

## RESULT 6

A41651  
retinoic acid receptor coregulator - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 20-Sep-1999  
C:Accession: A41651  
R:Yu, V.C.; Delbert, C.; Andersen, B.; Holloway, J.M.; Devary, O.V.; Naeae, A.M.; Ki  
Cell 67, 1251-1266, 1991  
A:Title: RXRbeta: a coregulator that enhances binding of retinoic acid, thyroid hormo  
A:Reference number: A41651; MUID:92103690  
A:Accession: A41651  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-451 <YUA>  
A:Cross-references: GB:M81766; NID:g206602; PIDN:AAA42025.1; PID:g206603  
C:Superfamily: unassigned erba-related proteins; erba transforming protein homology  
C:Keywords: zinc finger  
F:121-367/Domain: erba transforming protein homology <ERBA>

Query Match 72.4%; Score 55; DB 2; Length 451;  
Best Local Similarity 72.7%; Pred. No. 16;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 NAPPPLPPAP 12  
:|||||  
Db 36 SAPPPMPPPP 46

## RESULT 7

184718  
RXR-beta1 isoform - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 20-Aug-1999  
C:Accession: 184718  
R:Nagata, T.; Kanno, Y.; Ozato, K.; Taketo, M.  
Gene 142, 183-189, 1994  
A:Title: The mouse RXRb gene encoding RXR beta: genomic organization and two mRNA iso  
A:Reference number: 184718; MUID:94252565  
A:Accession: 184718  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-520 <RES>  
A:Cross-references: GB:D21831; NID:g497828; PIDN:BA04858.1; PID:g987669  
C:Genetics:  
A:Gene: RXRb  
A:Introns: 76/1; 148/3; 201/1; 362/1; 406/2; 437/1; 472/2  
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
F:190-436/Domain: erba transforming protein homology <ERBA>

Thu Jan 24 09:56:12 2002

us-09-485-529-103.rpr

Page 3

```
Query Match      72.4%; Score 55; DB 2; Length 520;
Best Local Similarity 72.7%; Pred. No. 18;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 NAPPPLPPAP 12
    :|||||:
Db 105 SAPPMPMP 115

RESULT 8
C69734
hypothetical protein FL5H1.18 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96734
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maitl, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C96734
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-138 <STO>
A:Cross-references: GB:AE005173; NID:g5902409; PIDN:AAD5511.1; GSPDB:GN00141
C:Genetics:
A:Gene: FL5H1.18
A:Map position: 1

Query Match      71.1%; Score 54; DB 2; Length 138;
Best Local Similarity 68.9%; Pred. No. 6.4;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12
    :|||||:
Db 71 PPPPLPPSP 79

RESULT 9
T30930
hypothetical protein D1007.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
R:Davidson, S.; Rohlfing, T.
submitted to the EMBL Data Library, August 1999
A:Description: The sequence of C. elegans cosmid D1007.
A:Reference number: Z20934
A:Accession: T30930
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-998 <DAV>
A:Cross-references: EMBL:AF003151; PIDN:AAB54225.1
C:Genetics:
A:Experimental source: strain Bristol N2; clone D1007
C:Genetics:
A:Map position: 1
A:introns: 54/3; 141/1; 314/2; 532/1; 711/3; 756/3; 907/2; 939/1
A:Note: D1007.7

Query Match      71.1%; Score 54; DB 2; Length 998;
Best Local Similarity 68.8%; Pred. No. 45;
Matches 11; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

OY 1 LNAPPPP---LPPAP 12
```

```
Db 714 LNAPPPGIRGPPAP 729
    :|||||:
RESULT 10
S60254
nuclear receptor co-repressor N-COR - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 08-Oct-1999
C:Accession: S60254
R:Hoerlein, A.J.; Naeae, A.M.; Heinzel, T.; Torchia, J.; Gloss, B.; Kurokawa, R.; Ry
Nature 377, 397-404, 1995
A:Title: Ligand-independent repression by the thyroid hormone receptor mediated by a
A:Reference number: S60254; MUID:96008539
A:Accession: S60254
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2453 <HOB>
A:Cross-references: EMBL:U35312; NID:g1022717; PIDN:AAB17125.1; PID:g1022718

Query Match      71.1%; Score 54; DB 2; Length 2453;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13
    :|||||:
Db 606 PPPPLPPPE 615

RESULT 11
I49729
HD protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C:Accession: I49729; I49730
R:Lin, B.; Nasir, J.; Macdonald, H.; Hutchinson, G.; Graham, R.; Rommens, J.M.; Hayde
Hum. Mol. Genet. 3, 85-92, 1994
A:Title: Sequence of the murine huntington disease gene: evidence for conservation an
A:Reference number: I49729; MUID:94214482
A:Accession: I49729
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3119 <RES>
A:Cross-references: GB:U23312; NID:g438804; PIDN:AAA37799.1; PID:g438805
A:Accession: I49730
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1521,2002-3119 <RE2>
A:Cross-references: GB:U23313; NID:g438806; PIDN:AAA37800.1; PID:g438807
C:Keywords: alternative splicing

Query Match      71.1%; Score 54; DB 2; Length 3119;
Best Local Similarity 81.8%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 APPPLPPAPQ 13
    :|||||:
Db 29 APPPPPPPPQ 39

RESULT 12
S32101
PHLP5A protein - common timothy (fragment)
C:Species: Phleum pratense (common timothy)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 03-Mar-1995
C:Accession: S32101
R:Butt, A.; Becker, W.M.; Petersen, A.; Schramm, G.; Schlaak, M.
submitted to the EMBL Data Library, February 1993
A:Description: Phl pv (timothy grass) major allergen bears at least two B-cell epitop
A:Reference number: S32101
A:Accession: S32101
```

A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-257 <BRU>  
 A:Cross-references: EMBL:X70942  
 C:Superfamily: grass pollen allergen IX

Query Match 69.7%; Score 53; DB 2; Length 257;  
 Best Local Similarity 81.8%; Pred. No. 15;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 PPPPLPPAPQ 13  
 |||||  
 Db 228 AKPPPLPPPPQ 238

## RESULT 13

A70856  
 Probable lppz protein - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: A70856  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: A70856  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-373 <COI>  
 A:Cross-references: GB:AL021287; GB:AL123456; NID:G3261508; PIDN:CAA16091.1; PID:el23775  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: lppz

Query Match 69.7%; Score 53; DB 2; Length 373;  
 Best Local Similarity 88.9%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
 |||||  
 Db 51 PPPPLPPVP 59

## RESULT 14

T19721  
 hypothetical protein C34E11.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1989 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
 C:Accession: T19721  
 R:McMurray, A.  
 submitted to the EMBL Data Library, November 1995  
 A:Reference number: Z19169  
 A:Accession: T19721  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-731 <WIL>  
 A:Cross-references: EMBL:Z67754; PIDN:CAA91752.1; GSPDB:GN00028; CESP:C34E11.2  
 A:Experimental source: clone C34E11  
 C:Genetics:  
 A:Gene: CESP:C34E11.2  
 A:Map position: X  
 A:Introns: 54/3; 149/3; 183/1; 224/2; 273/2; 330/3; 378/3; 416/1; 510/1; 578/2; 616/2  
 C:Superfamily: Caenorhabditis elegans hypothetical protein C34E11.2

Query Match 69.7%; Score 53; DB 2; Length 731;  
 Best Local Similarity 61.5%; Pred. No. 43;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 LNAPPPPLPPAPQ 13  
 ::|||  
 Db 494 MSSPPPLPPRPQ 506

## RESULT 15

T32889  
 hypothetical protein C34B2.9 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T32889  
 R:Graves, T.; Suterer, C.; Hawkins, M.; Wilson, R.  
 submitted to the EMBL Data Library, January 1998  
 A:Description: The sequence of C. elegans cosmid C34B2.  
 A:Reference number: Z21241  
 A:Accession: T32889  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-238 <GRA>  
 A:Cross-references: EMBL:AF043693; PIDN:AAB97541.1; GSPDB:GN00019; CESP:C34B2.9  
 A:Experimental source: strain Bristol N2; clone C34B2  
 C:Genetics:  
 A:Gene: CESP:C34B2.9  
 A:Map position: 1  
 A:Introns: 35/3; 91/2; 200/2

Query Match 68.4%; Score 52; DB 2; Length 238;  
 Best Local Similarity 88.9%; Pred. No. 18;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
 |||||  
 Db 132 PPPPLPPAP 140

Search completed: January 22, 2002, 16:24:48  
 Job time: 68 sec

Thu Jan 24 09:56:12 2002

us-09-485-529-103.rpr

Page 5





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:30:28 ; Search time 17.09 Seconds  
(without alignments)  
27.890 Million cell updates/sec

Title: US-09-485-529-103  
Perfect score: 76  
Sequence: 1 LNAAPPPLPPAPQ 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SWISSProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	75.0	282	1	ATF5_HUMAN
2	57	75.0	901	1	Y298_HUMAN
3	55	72.4	367	1	CD55_HUMAN
4	55	72.4	369	1	CD55_MOUSE
5	55	72.4	458	1	RXR_B_RAT
6	55	72.4	520	1	RXR_B_MOUSE
7	54	71.1	591	1	NMT_MOUSE
8	54	71.1	2440	1	NCRL_HUMAN
9	54	71.1	2453	1	NCRL_MOUSE
10	54	71.1	3119	1	HD_MOUSE
11	53	69.7	724	1	CCT1_MOUSE
12	53	69.7	726	1	CCT1_HUMAN
13	53	69.7	727	1	CCT1_HORSE
14	52	68.4	281	1	PASL_HUMAN
15	52	68.4	288	1	SMN_MOUSE
16	52	68.4	389	1	NDEP_MOUSE
17	52	68.4	408	1	G3PT_HUMAN
18	52	68.4	543	1	VP61_NPVAC
19	52	68.4	736	1	DVL2_MOUSE
20	52	68.4	809	1	CN4D_HUMAN
21	52	68.4	1132	1	BAT3_HUMAN
22	52	68.4	1164	1	PHOS_HUMAN
23	52	68.4	1733	1	VNOA_PPRKA
24	51	67.1	223	1	TOMB_XANCP
25	51	67.1	294	1	SMN_HUMAN
26	51	67.1	311	1	CDX2_HUMAN
27	51	67.1	372	1	PEXC_ARATH
28	51	67.1	399	1	HXA_HUMAN
29	51	67.1	744	1	IE63_HCVAC
30	51	67.1	808	1	Y066_NPVAC
31	51	67.1	825	1	RCAL1_YEAST
32	51	67.1	884	1	ANDR_EULFC
33	51	67.1	2911	1	FBN2_HUMAN

## ALIGNMENTS

RESULT	ID	ATF5_HUMAN	STANDARD	PRT	282 AA.
34	50	65.8	165	1	Y084_RICPR
35	50	65.8	288	1	FXE3_MOUSE
36	50	65.8	307	1	RX33_CAELT
37	50	65.8	309	1	HXA4_CHICK
38	50	65.8	401	1	HME1_MOUSE
39	50	65.8	415	1	ACRO_PIG
40	50	65.8	440	1	G3PT_MOUSE
41	50	65.8	487	1	EBN2_EBV
42	50	65.8	533	1	RXR_B_HUMAN
43	50	65.8	559	1	WAS1_HUMAN
44	50	65.8	601	1	3BP1_MOUSE
45	50	65.8	736	1	DVL2_HUMAN

09263 rickettsia	09414 mus musculu	020365 caenorhabdi	P17277 gallus galli	P09065 mus musculu	P08001 sus scrofa	O64467 mus musculu	P12978 epstein-bar	P28702 homo sapien	O92558 homo sapien	P55194 mus musculu	O14641 homo sapien
------------------	-------------------	--------------------	---------------------	--------------------	-------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------

ATF5_HUMAN	STANDARD	PRT	282 AA.
AC	Q912D1; Q9UNQ3;		
DT	20-AUG-2001 (Rel. 40, Created)		
DT	20-AUG-2001 (Rel. 40, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	CYCLOC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-5 (ACTIVATING		
DE	TRANSCRIPTION FACTOR 5) (TRANSCRIPTION FACTOR ATF5).		
GN	ATF5 OR ATF5.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20558615; PubMed=11087824;		
RA	White J.H., McIlhinney R.A.J., Wise A., Ciruela F., Chan W.Y.,		
RA	Emson P.C., Billington A., Marshall F.H.;		
RT	"The GABAB receptor interacts directly with the related transcription		
RT	factors CREB2 and ATF5."		
RT	Proc. Natl. Acad. Sci. U.S.A. 97:13967-13972(2000).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RA	Kohroki J., Tanaka K.;		
RL	Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.		
RN	[3]		
RP	SEQUENCE OF 161-282 FROM N.A.		
RX	MEDLINE=9303793; PubMed=10373550;		
RA	Patl D., Weistrich M.L., Pion S.E.;		
RT	"Human Cdc34 and Rad6b ubiquitin-conjugating enzymes target repressors		
RT	of cyclic AMP-induced transcription for proteolysis."		
RL	Mol. Cell. Biol. 19:3001-3013(1999).		
CC	- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE)		
CC	(CONSENSUS: 5'GTGACGT(A/C)(A/G)-3'), A SEQUENCE PRESENT IN MANY		
CC	VIRAL AND CELLULAR PROMOTERS.		
CC	- SUBUNIT: BINDS DNA AS A DIMER (POTENTIAL). INTERACTS WITH GABAB		
CC	RECEPTOR.		
CC	- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).		
CC	- SIMILARITY: TO OTHER BZIP PROTEINS.		
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
DR	EMBL; AF305687; AAG2558.1; -		
DR	EMBL; AF021663; BA878477.2; -		
DR	EMBL; AF101388; AAD28370.1; -		
DR	InterPro; IPR002637; Ham1p-like.		
DR	InterPro; IPR002965; P-rich_extensin.		
DR	InterPro; IPR001871; bzip.		
DR	Pfam; PF00170; bzip; 1.		

```

DR PRINTS: PRO1217; PRICHEXTENSN.
DR PRODOM: PD004952; Hampl_1like; 1.
DR SMART: SMO0338; BRIZ; 1.
DR PROSITE: PS00036; BRIP_BASIC; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW Multigene family.
FT DOMAIN 123 139 POLY-PRO.
FT DOMAIN 186 194 POLY-PRO.
FT DNA_BIND 210 230 BASIC MOTIF.
FT DOMAIN 236 250 LEUCINE-ZIPPER (PROBABLE).
FT CONFLICT 161 163 LLA -> RHE (IN REF. 3).
SQ SEQUENCE 282 AA; 30674 MW; DDB2P90/GCA0215A0.CRC64;

Query Match
Best Local Similarity 75.0%; Score 57; DB 1; Length 282;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12
DB 131 PPPPLPPAP 139

RESULT 2
Y298_HUMAN STANDARD; PRT; 901 AA.
ID Y298_HUMAN
AC 015016;
DT 15-JUL-1998 (Rel. 36, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0298 (FRAGMENT).
GN KIAA0298.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;

RN SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Nagase T., Ishikawa K.-I., Seki N., Nakajima D., Ohira M.,
RA Miyajima N., Kotani H., Nomura N., Ohara O.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN SEQUENCE OF 305-901 FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB002296; BAA20758.2; -
DR PROSITE: PS50119; ZF_BBOX; 2.
KW Hypothetical protein; Zinc-finger; Repeat.
FT NON_TER 1
FT ZN_FING 33 79 B BOX-TYPE 1.
FT ZN_FING 93 134 B BOX-TYPE 2.
FT DOMAIN 486 532 PRO-RICH.
SQ SEQUENCE 901 AA; 99510 MW; 8A5C91B014F5F556.CRC64;

Query Match
Best Local Similarity 75.0%; Score 57; DB 1; Length 901;
Matches 90.0%; Pred. No. 15;

```

```

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAPQ 13
DB 500 PPPPLPPAPQ 509

RESULT 3
CD55_HUMAN STANDARD; PRT; 367 AA.
ID CD55_HUMAN
AC Q13319;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYCLIN-DEPENDENT KINASE 5 ACTIVATOR 2 PRECURSOR (CDK5 ACTIVATOR 2)
DE (CYCLIN-DEPENDENT KINASE 5 REGULATORY SUBUNIT 2) (P39) (P39I).
GN CDK5R2 OR NCK5A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;

RN SEQUENCE FROM N.A.
RC TISSUE=Hippocampus;
RC MEDLINE=96070784; PubMed=7592934;
RA Tang D., Yeung J., Lee K.-Y., Matsushita M., Matsui H., Tomizawa K.,
RA Hatase O., Wang J.H.;
RT "An isoform of the neuronal cyclin-dependent kinase 5 (cdk5)
RT activator.";
RL J. Biol. Chem. 270:26897-26903(1995).
CC -1 FUNCTION: ACTIVATOR OF CDK5/TPKII.
CC -1 SUBUNIT: HETERODIMER OF A CATALYTIC SUBUNIT AND A REGULATORY
CC SUBUNIT.
CC -1 TISSUE SPECIFICITY: BRAIN AND NEURON SPECIFIC.
CC -1 SIMILARITY: BELONGS TO THE CYCLIN-DEPENDENT KINASE 5 ACTIVATOR
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U34051; AAC50278.1; -
DR MIM: 603764; -
FT PROPEP 1 ? POTENTIAL.
FT CHAIN ? 367 CYCLIN-DEPENDENT KINASE 5 ACTIVATOR 2.
FT DOMAIN 75 78 POLY-LYS.
FT DOMAIN 130 133 POLY-ALA.
FT DOMAIN 155 163 POLY-PRO.
SQ SEQUENCE 367 AA; 38704 MW; D8CB7CB8E0A8200.CRC64;

Query Match
Best Local Similarity 72.4%; Score 55; DB 1; Length 367;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAPQ 13
DB 155 PPPPLPPAPQ 164

RESULT 4
CD55_MOUSE STANDARD; PRT; 369 AA.
ID CD55_MOUSE
AC O35926; O35277;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CYCLIN-DEPENDENT KINASE 5 ACTIVATOR 2 PRECURSOR (CDK5 ACTIVATOR 2)
DE (CYCLIN-DEPENDENT KINASE 5 REGULATORY SUBUNIT 2) (P39) (P39I).

```

```

GN CDR5R2 OR NCR5A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX MEDLINE=98322126; PubMed=9655938;
RA Nilsen F., Baeckstrom A., Bark C.;
RT "Molecular cloning and characterization of a mouse gene encoding an
RT isoform of the neuronal cyclin-dependent kinase 5 (CDK5) activator.";
RL Biochim. Biophys. Acta 1398:371-376(1998).
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RA Zheng M., Lemm R.K.H.;
RT "Comparative analysis of gene expression of the cyclin-dependent
RT kinase 5 (cdk5) activators p35 and p39 in the rat CNS and embryonic
RT development revealed distinctive overlap with cdk5.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACTIVATOR OF CDK5/TPK1T.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC SUBUNIT AND A REGULATORY
CC SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE CYCLIN-DEPENDENT KINASE 5 ACTIVATOR
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U90267; AAC53595.1;
DR EMBL: AF016393; AAB69709.1;
DR MGI: MGI:130828; Cdk5r2.
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PR01217; PRICHEXTENS.
DR PROEP 1
FT CHAIN ? 369
FT DOMAIN ? 78 CYCLIN-DEPENDENT KINASE 5 ACTIVATOR 2.
FT DOMAIN 155 163 POLY-LYS.
FT DOMAIN 348 351 POLY-PRO.
FT CONFLICT 91 91 G->V (IN REF. 2).
FT CONFLICT 172 172 S->G (IN REF. 2).
FT CONFLICT 191 191 R->G (IN REF. 2).
FT CONFLICT 197 197 S->R (IN REF. 2).
FT CONFLICT 220 220 L->W (IN REF. 2).
FT CONFLICT 248 248 S->R (IN REF. 2).
FT CONFLICT 253 253 A->G (IN REF. 2).
SQ SEQUENCE 369 AA; 38914 MW; DA4EB1FD4D48CE66 CRC64;

Query Match 72.4%; Score 55; DB 1; Length 369;
Best Local Similarity 90.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13
DB 155 PPPPPPPAPQ 164

```

```

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RX MEDLINE=92103690; PubMed=1662118;
RA Yu V.C., Delser C., Andersen B., Holloway J.M., Devary O.V.,
RA Naess A.M., Kim S.Y., Boutin J.M., Glass C.K., Rosenfeld M.G.;
RT "RXR beta: a coregulator that enhances binding of retinoic acid,
RT thyroid hormone, and vitamin D receptors to their cognate response
RT elements.";
RL Cell 67:1251-1266(1991).
CC -1- FUNCTION: INVOLVED IN RETINOIC ACID RESPONSE PATHWAY. BINDS
CC 9-CIS RETINOIC ACID (9C-RA).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M81766; AAA42025.1; ALT_INIT.
DR HSSP: P19793; 2NLT.
DR InterPro: IPR000536; Hormone_rec_1lg.
DR InterPro: IPR001628; zf-C4.
DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00105; zf-C4; 1.
DR SMART: SM00430; HOL1; 1.
DR SMART: SM00398; zNF_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW zinc-finger; Multigene family.
FT FT 1 1
FT DOMAIN <1 129 MODULATING (BY SIMILARITY).
FT DNA_BIND 130 195 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 130 150 C4-TYPE.
FT ZN_FING 166 190 C4-TYPE.
FT DOMAIN 196 255 HINGE.
FT DOMAIN 256 458 LIGAND-BINDING (BY SIMILARITY).
SQ SEQUENCE 458 AA; 49078 MW; 22D11956B11DB4BE CRC64;

Query Match 72.4%; Score 55; DB 1; Length 458;
Best Local Similarity 72.7%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 NAPPPLPPAP 12
DB 43 SAPPPLPPPP 53

RESULT 6
RXRB_MOUSE
ID RXRB_MOUSE STANDARD: PRT; 520 AA.
AC P28704; P33243;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RETINOIC ACID RECEPTOR RXR-BETA (MHC CLASS I REGULATORY ELEMENT
DE BINDING PROTEIN H-2RIIBP).
GN RXRB OR NR2B2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```



TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL: U77356; AAB8687.1; -;  
 CC EMBL: Y07609; CAA68878.1; -;  
 CC MGD: MGI:109150; Mat.  
 CC InterPro: IPR003015; HLH\_MYC.  
 CC InterPro: IPR001092; HLH\_dim.  
 CC Pfam: PF00010; HLH; 1.  
 CC SMART: SM00353; HLH; 1.  
 CC PROSITE: PS00038; HELIX\_LOOP\_HELIX; FALSE\_NEG.  
 CC Transcription regulation; Repressor; Nuclear protein; DNA-binding.  
 CC DNA\_BIND 224 235 BASIC MOTIF (POTENTIAL) (BY SIMILARITY).  
 CC DOMAIN 236 272 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 CC DOMAIN 273 301 LEUCINE-ZIPPER (POTENTIAL).  
 CC CONFLICT 379 379 T -> P (IN REF. 2).  
 CC CONFLICT 392 392 V -> A (IN REF. 2).  
 CC CONFLICT 402 403 EE -> OO (IN REF. 2).  
 CC CONFLICT 414 414 G -> A (IN REF. 2).  
 CC CONFLICT 431 431 V -> A (IN REF. 2).  
 CC CONFLICT 465 465 A -> P (IN REF. 2).  
 CC CONFLICT 525 525 T -> A (IN REF. 2).  
 CC CONFLICT 558 558 G -> A (IN REF. 2).  
 CC SEQUENCE 591 AA; 63311 MW; 3FAE61A10CF4142F CRC64;

Query Match 71.1%; Score 54; DB 1; Length 591;  
 Best Local Similarity 90.0%; Pred. No. 21;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAPQ 13  
 DB 105 PPPPLPPAPQ 114

-----

RESULT 8  
 NCRL\_HUMAN STANDARD; PRT; 2440 AA.  
 AC 075376; Q9UPV5; Q90218;  
 DT 20-AUG-2001 (Rel. 40; Created)  
 DT 20-AUG-2001 (Rel. 40; Last sequence update)  
 DT 20-AUG-2001 (Rel. 40; Last annotation update)  
 DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR).  
 GN NCOR1 OR KIAA1047.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP TISSUE: Fetal brain;  
 RX MEDLINE=96393736; PubMed=9724795;  
 RA Wang J., Hoshino T., Redner R.L., Kajigaya S., Liu J.M.;  
 RT "ETO, fusion partner in t(8;21) acute myeloid leukemia, represses  
 RT transcription by interaction with the human N-COR/MSI3/HDAC1  
 RT complex.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10860-10865(1998).  
 RN [2]  
 RP SEQUENCE OF 782-2440 FROM N.A.  
 RC TISSUE: Brain;  
 RX MEDLINE=99397452; PubMed=10470851;  
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohata O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIV.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";

-----

RL DNA Res. 6:197-205(1999).  
 RN [3]  
 RP SEQUENCE OF 974-2440 FROM N.A.  
 RX MEDLINE=99375328; PubMed=10444336;  
 RA Nagaya T., Chen K.-S., Fujieda M., Ohmori S., Richer J.K.,  
 RA Horvitz K.B., Lupski J.R., Seo H.;  
 RT "Localization of the human nuclear receptor co-repressor (hn-Cor) gene  
 RT between the CMT1a and the SMS critical regions of chromosome  
 RT 17p11.2.";  
 RL Genomics 59:339-341(1999).  
 CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME  
 CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS  
 CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.  
 CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B  
 CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES  
 CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE  
 CC ABSENCE OF LIGAND.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
 CC DOMAINS (ID1 AND ID2).  
 CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
 CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND  
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RAR. SEQUENCES  
 CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
 CC SPECIFICITY.  
 CC -1- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A+SANT-B).  
 CC -1- SIMILARITY: CONTAINS 2 CORNR BOX.  
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
 CC FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL: AF044209; AAC33550.1; -;  
 CC EMBL: AB028970; BAA82999.1; -;  
 CC EMBL: AB019524; BAA75814.1; -;  
 CC MIM: 600849; -;  
 CC InterPro: IPR001005; MYB\_DNA\_bind.  
 CC Pfam: PF00249; MYB\_DNA-binding; 2.  
 CC SMART: SM00395; SANT; 2.  
 CC PROSITE: PS50090; MYB\_3; 1.  
 CC Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
 CC Coiled coil.  
 CC FT 174 216 COILED COIL (POTENTIAL).  
 CC FT- DOMAIN 254 312 INTERACTION WITH SIN3A/B.  
 CC FT- DOMAIN 299 328 COILED COIL (POTENTIAL).  
 CC FT- DOMAIN 437 482 SANT-A (POTENTIAL).  
 CC FT- DNA\_BIND 625 670 SANT-B (POTENTIAL).  
 CC FT- DOMAIN 501 557 COILED COIL (POTENTIAL).  
 CC FT- DOMAIN 607 617 PRO-RICH.  
 CC FT- DOMAIN 988 1816 INTERACTION WITH ETO.  
 CC FT- DOMAIN 2055 2059 CORNR BOX OF ID1.  
 CC FT- DOMAIN 2263 2267 CORNR BOX OF ID2.  
 CC FT- DOMAIN 58 64 POLY-GLN.  
 CC FT- DOMAIN 593 603 POLY-ALA.  
 CC FT- DOMAIN 1032 1035 POLY-PRO.  
 CC FT- DOMAIN 1707 1712 POLY-ALA.  
 CC FT- DOMAIN 1952 1963 POLY-SER.  
 CC FT- CONFLICT 1014 1014 L -> V (IN REF. 2).  
 CC FT- CONFLICT 1508 1509 PP -> SS (IN REF. 2).  
 CC FT- CONFLICT 1561 1561 W -> R (IN REF. 2).  
 CC FT- CONFLICT 1567 1567 Q -> H (IN REF. 2).  
 CC SEQUENCE 2440 AA; 270263 MW; 60A4D7964D00EDAB CRC64;

```

Query Match      71.1%; Score 54; DB 1; Length 2440;
Best Local Similarity 80.0%; Pred. No. 82;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      4 PPPPPAPQ 13
      ||||| 1:
Db      607 PPPPPPEPE 616

RESULT 9
NCRL_MOUSE
ID NCRL_MOUSE STANDARD; PRT; 2453 AA.
AC Q60974; Q60812;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR) (RETINOID X RECEPTOR
DE INTERACTING PROTEIN 13) (RIP13).
CN NCOR1 OR RXRIP13.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN 11
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC TISSUE-Pituitary;
RX MEDLINE=96008539; PubMed=7566114;
RA Hoerlein A.J., Naeae A.M., Heinzel T., Torchia J., Gloss B.,
RA Kurokawa R., Ryan A., Kamel Y., Soederstrom M., Glass C.K.,
RA Rosenfeld M.G.;
RT "Ligand-independent repression by the thyroid hormone receptor
RT mediated by a nuclear receptor co-repressor.";
RL Nature 377:397-404(1995).
[2]
RP SEQUENCE OF 1792-2453 FROM N.A. (ISOFORM LONG).
RC TISSUE-Liver;
RX MEDLINE=95280959; PubMed=7760852;
RA Seol W., Choi H.S., Moore D.D.;
RT "Isolation of proteins that interact specifically with the retinoid X
RT receptor: two novel orphan receptors.";
RL Mol. Endocrinol. 9:72-85(1995).
CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
CC AND HISTONE DEACETYLASE HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
CC WITH THE THYROID (TR) AND THE RETINOID ACID RECEPTORS (RAR) IN THE
CC ABSENCE OF LIGAND.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: OBIDUOUS.
CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
CC DOMAINS (ID1 AND ID2).
CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND
CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES
CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR
CC SPECIFICITY.
CC -1- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A+SANT-B).
CC -1- SIMILARITY: CONTAINS 2 CORNR BOX.
CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; U35312; AAB17125.1; -
DR EMBL; U22016; AAC52168.1; -
DR MGI; MGI:1349717; Ncor1.
DR InterPro: IPR001005; Myb_DNA_bind.
DR Pfam; PF00248; myb_DNA-binding; 2.
DR SMART; SM00395; SANT; 2.
DR PROSITE; PS50090; MYB_3; 1.
KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;
KW Coiled coil; Alternative splicing.
FT DOMAIN 174 216
FT 254 312 COILED COIL (POTENTIAL).
FT 299 328 INTERACTION WITH SIN3A/B (BY SIMILARITY).
FT 437 482 COILED COIL (POTENTIAL).
FT 624 669 SANT-A (POTENTIAL).
FT 501 550 SANT-B (POTENTIAL).
FT 606 616 COILED COIL (POTENTIAL).
FT 2073 2077 PRO-RICH.
FT 2277 2281 CORNR BOX OF ID1.
FT 58 64 CORNR BOX OF ID2.
FT 593 602 POLY-GLN.
FT 1044 1047 POLY-ALA.
FT 1713 1718 POLY-PRO.
FT 1968 1979 POLY-SER.
FT 2333 2371 POLY-SER.
FT 1952 1952 MISSING (IN SHORT ISOFORM).
FT 2090 2090 I -> T (IN REF. 2).
FT 2090 2090 A -> P (IN REF. 2).
SQ SEQUENCE 2453 AA; 270640 MW; 52208B4032F7E6A CRC64;

Query Match      71.1%; Score 54; DB 1; Length 2453;
Best Local Similarity 80.0%; Pred. No. 82;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      4 PPPPPAPQ 13
      ||||| 1:
Db      606 PPPPPPEPE 615

RESULT 10
NCRL_MOUSE
ID NCRL_MOUSE STANDARD; PRT; 3119 AA.
AC P42859;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HUNTINGTIN (HUNTINGTON'S DISEASE PROTEIN HOMOLOG) (HD PROTEIN).
CN HD OR HDH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6; TISSUE-Brain, and Splen;
RX MEDLINE=94214482; PubMed=8162057;
RA Lin B., Nasir J., Macdonald H., Hutchinson G., Graham R.K.,
RA Rommens J.M., Hayden M.R.;
RT "Sequence of the murine Huntington disease gene: evidence for
RT conservation, alternate splicing and polymorphism in a triplet
RT repeat.";
RL Hum. Mol. Genet. 3:85-92(1994).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94278649; PubMed=8009370;
RA Barnes G.T., Duyao M.P., Ambrose C.M., McNeil S., Persichetti F.,
RA Strinchi J., Gusella J.F., Macdonald M.E.;
RT "Mouse Huntington's disease gene homolog (Hdh).";
RL Somat. Cell Mol. Genet. 20:87-97(1994).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95375771; PubMed=7647777;

```

RA Trotter Y., Devys D., Imbert G., Saudou F., An I., Lutz Y.,  
 RA Weber C., Agid Y., Hirsch E.C., Mandel J.L.;  
 RT "Cellular localization of the Huntington's disease protein and  
 RT discrimination of the normal and mutated form.";  
 RL Nat. Genet. 10:104-110(1995).  
 [4]  
 RN SEQUENCE OF 1-181 FROM N.A.  
 RP MEDLINE=95278941; PubMed=7759106;  
 RX Lin B., Nasir J., Kalchman M.A., McDonald H., Zeisler J.,  
 RA Goldberg Y.P., Hayden M.R.;  
 RT "Structural analysis of the 5' region of mouse and human Huntington  
 RT disease genes reveals conservation of putative promoter region and  
 RT di- and trinucleotide polymorphisms.";  
 RL Genomics 25:707-715(1995).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN MICROTUBULE-MEDIATED TRANSPORT OR  
 CC VESICLE FUNCTION.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- ALTERNATIVE PRODUCTS: ALTERNATIVE SPLICING SEEMS TO GIVE RISE TO A  
 CC 480 AMINO ACIDS SHORTER ISOFORM. THIS PRODUCT CANNOT BE EXPLAINED  
 CC BY A SIMPLE SPLICING EVENT.  
 CC -1- TISSUE SPECIFICITY: THE HIGHEST LEVEL IS SEEN THROUGHOUT THE  
 CC BRAIN, BUT IT IS ALSO FOUND IN THE STOMACH, HEART, TESTIS, ADIPOSE  
 CC TISSUE, MUSCLE, SPLEEN, LIVER, AND KIDNEY.  
 CC -1- DEVELOPMENTAL STAGE: PREDOMINANT EXPRESSION IN NEURONAL TISSUES AT  
 CC ALL DEVELOPMENTAL STAGES. IN 14.5 DAY OLD EMBRYOS, IT IS ALSO  
 CC DETECTED IN NON-NEURONAL TISSUES. THIS EXPRESSION IS DOWN-  
 CC REGULATED IN LATER STAGES OF DEVELOPMENT.  
 CC -1- POLYMORPHISM: THE FIRST POLY-PRO REPEAT STRETCH DIFFERS IN LENGTH  
 CC BY ONE UNIT (THREE) IN MUS SPRETUS STRAIN COMPARED TO OTHER  
 CC STRAINS (FOUR). THE POLY-GLN REGION DOES NOT APPEAR TO BE  
 CC POLYMORPHIC, EXPLAINING THE ABSENCE OF A MURINE HD-LINE DISORDER.  
 CC -1- SIMILARITY: CONTAINS 10 HEAT REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE HUNTINGTIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L23312; AAA37799.1; -;  
 DR EMBL; L23313; AAA37800.1; -;  
 DR EMBL; L28827; AAA89100.1; ALT. SEQ.  
 DR EMBL; U24233; AAC52218.1; -;  
 DR EMBL; L34024; AAA91085.1; -;  
 DR EMBL; L34008; AAA91085.1; JOINED.  
 DR EMBL; L34021; AAA91085.1; JOINED.  
 DR EMBL; L34022; AAA91085.1; JOINED.  
 DR EMBL; L34023; AAA91085.1; JOINED.  
 DR EMBL; MGI:96067; Hdh.  
 DR MGI:96067; Hdh.  
 DR InterPro: IPR000091; Huntingtin.  
 DR PRINTS: PR00375; HUNTINGTIN.  
 DR Alternative splicing; Polymorphism; Repeat.  
 KW DOMAIN 182 306 HEAT REPEATS DOMAIN 1.  
 FT DOMAIN 723 920 HEAT REPEATS DOMAIN 2.  
 FT DOMAIN 1512 1553 HEAT REPEATS DOMAIN 3.  
 FT DOMAIN 18 24 POLY-GLN.  
 FT DOMAIN 25 45 POLY-PRO.  
 FT DOMAIN 49 59 POLY-PRO.  
 FT DOMAIN 1417 1420 POLY-THR.  
 FT DOMAIN 1696 1699 POLY-GLU.  
 FT DOMAIN 2615 2620 POLY-GLU.  
 FT VARSLIC 1522 2001 MISSING (IN SHORT ISOFORM).  
 FT CONFLICT 2 2 A -> G (IN REF. 1 AND 4).  
 FT CONFLICT 29 29 A -> P (IN REF. 2).  
 FT CONFLICT 116 116 M -> L (IN REF. 2 AND 4).  
 FT CONFLICT 138 138 N -> D (IN REF. 1).  
 FT CONFLICT 521 521 S -> P (IN REF. 1).  
 FT CONFLICT 524 524 A -> P (IN REF. 1).  
 FT CONFLICT 533 533 A -> P (IN REF. 1).  
 FT CONFLICT 607 607 A -> T (IN REF. 1).

FT CONFLICT 769 769 D -> E (IN REF. 2).  
 FT CONFLICT 972 972 S -> R (IN REF. 1).  
 FT CONFLICT 1106 1106 W -> C (IN REF. 1).  
 FT CONFLICT 1240 1240 T -> N (IN REF. 1).  
 FT CONFLICT 1384 1384 N -> T (IN REF. 1).  
 FT CONFLICT 1827 1827 H -> Y (IN REF. 1).  
 FT CONFLICT 1979 1980 H -> Y (IN REF. 1).  
 FT CONFLICT 2062 2062 PF -> SS (IN REF. 1).  
 FT CONFLICT 2570 2570 D -> G (IN REF. 1).  
 FT CONFLICT 2570 2570 S -> N (IN REF. 1).  
 FT CONFLICT 2866 2866 E -> V (IN REF. 1).  
 FT CONFLICT 2877 2877 V -> G (IN REF. 1).  
 FT CONFLICT 2882 2882 Q -> G (IN REF. 1).  
 FT CONFLICT 2887 2887 O -> H (IN REF. 1).  
 FT CONFLICT 2915 2915 A -> T (IN REF. 1).  
 FT CONFLICT 3025 3025 P -> S (IN REF. 3).  
 FT CONFLICT 3062 3063 QV -> LM (IN REF. 1).  
 FT CONFLICT 3095 3096 VV -> EE (IN REF. 1).  
 SQ SEQUENCE 3119 AA; 344685 MW; ECA42B5916F50F4F CRC64;  
 Query Match 71.1%; Score 54; DB 1; Length 3119;  
 Best Local Similarity 81.8%; Pred. No. 1e+02;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 3 APPPPPPAPQ 13  
 DB 29 APPPPPPPPQ 39  
 RESULT 11  
 ID CCTL\_MOUSE STANDARD; PRT; 724 AA.  
 AC Q9QWV9; Q920U7;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CYCLIN T1 (CYCLIN T) (CYCT1).  
 GN CCNT1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C;  
 RA MEDLINE=99059742; PubMed=9843510;  
 RX Bieniasz P.D., Girdina T.A., Bogerd H.P., Cullen B.R.;  
 RT "Recruitment of a protein complex containing Tat and cyclin T1 to TAR  
 RT governs the species specificity of HIV-1 Tat.";  
 RL EMBO J. 17:7056-7065(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=99051315; PubMed=9832504;  
 RA Gardner M.E., Wei P., KewalRamani V.N., Mayall T.P., Herrmann C.H.,  
 RA Rice A.P., Littman D.R., Jones K.A.;  
 RT "The interaction between HIV-1 Tat and human cyclin T1 requires zinc  
 RT and a critical cysteine residue that is not conserved in the murine  
 RT Cycl1 protein.";  
 RL Genes Dev. 12:3512-3527(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
 RC TISSUE=Spleen;  
 RX MEDLINE=99263519; PubMed=10329126;  
 RA Kwak Y.T., Ivanov D., Guo J., Nee E., Gaynor R.B.;  
 RT "Role of the human and murine cyclin T proteins in regulating HIV-1  
 RT Tat-activation.";  
 RL J. Mol. Biol. 288:57-69(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=99145542; PubMed=9990016;  
 RA Fujinaga K., Taube R., Wimmer J., Cujec T.P., Peterlin B.M.;

"Interactions between human cyclin T, Tat, and the transactivation response element (TAR) are disrupted by a cysteine to tyrosine substitution found in mouse cyclin T.";  
Proc. Natl. Acad. Sci. U.S.A. 96:1285-1290(1999).  
-1- FUNCTION: REGULATORY SUBUNIT OF THE CYCLIN-DEPENDENT KINASE PAIR (CDK9/CYCLIN T) COMPLEX, ALSO CALLED POSITIVE TRANSCRIPTION ELONGATION FACTOR B (P-TEFB), WHICH IS PROPOSED TO FACILITATE THE TRANSITION FROM ABORTIVE TO PRODUCTION ELONGATION BY PHOSPHORYLATING THE CTD (CARBOXY-TERMINAL DOMAIN) OF THE LARGE SUBUNIT OF RNA POLYMERASE II (RNAP II). DOES NOT BIND EFFICIENTLY TO THE TRANSCRIPTION DOMAIN OF THE HIV-1 NUCLEAR TRANSCRIPTIONAL ACTIVATOR, TAT.  
-1- SUBUNIT: ASSOCIATES WITH CDK9 TO FORM P-TEFB. CYCLIN T1 IS THE PREDOMINANT CYCLIN ASSOCIATED WITH CDK9.  
-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
-1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. STRONG TO CYCLIN C SUBFAMILY.  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF095640; AAD13656.1; -  
DR EMBL; AF109179; AAD19654.1; -  
DR EMBL; AF087662; AAD17798.1; -  
DR EMBL; AF113951; AAD17205.1; -  
DR MGD; MGI:1328363; Ccml1.  
DR InterPro; IPR000553; Cyclin.  
DR SMART; SM00385; CYCLIN; 1.  
DR PROSITE; PS00292; CYCLINS; FALSE\_NEG.  
KW Cyclin; Cell cycle; Cell division; Coiled coil;  
KW Transcription regulation; Nuclear protein.  
FT DOMAIN 253 270 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 254 272 TAT:TAR RECOGNITION MOTIF (TRM).  
FT DOMAIN 319 322 POLY-SER.  
FT DOMAIN 384 425 COILED COIL (POTENTIAL).  
FT DOMAIN 516 525 POLY-HIS.  
FT DOMAIN 565 569 POLY-SER.  
FT DOMAIN 715 723 POLY-PRO.  
FT CONFLICT 48 48 V -> M (IN REF. 3).  
SQ SEQUENCE 724 AA; 80565 MW; 7396E428F5A8B91B CRC64;  
  
Query Match 69.7%; Score 53; DB 1; Length 724;  
Best Local Similarity 80.0%; Pred. No. 33;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 PPPPLPPAPQ 13  
DB 715 PPPPLPPLPK 724  
  
RESULT 12  
CCT1\_HUMAN STANDARD; PRT; 726 AA.  
AC 060563; 060581;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CYCLIN T1 (CYCLIN T) (CCT1).  
GN CCT1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_taxid=9606;  
RX [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=T-cell lymphoma;  
ME MEDLINE=96150851; Pubmed=9491887;

RA Wei P., Garber M.E., Fang S.-M., Fischer W.H., Jones K.A.;  
RT "A novel CDK9-associated C-type cyclin interacts directly with HIV-1  
RT Tat and mediates its high-affinity, loop-specific binding to TAR  
RT RNA.";  
RT Cell 92:451-462(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast cancer;  
RX MEDLINE=98167917; Pubmed=9499409;  
RA Peng J.-M., Zhu Y., Milton J.T., Price D.H.;  
RT "Identification of multiple cyclin subunits of human P-TEFB.";  
RT Genes Dev. 12:755-762(1998).  
RN [3]  
RP FUNCTION.  
RX MEDLINE=99263518; Pubmed=10329125;  
RA Ivanov D., Kwak Y.T., Nee E., Guo J., Garcia-Martinez L.F.,  
RA Gaynor R.B.;  
RT "Cyclin T1 domains involved in complex formation with Tat and TAR RNA  
RT are critical for tat-activation.";  
RT J. Mol. Biol. 288:41-56(1999).  
RN [4]  
RP MUTAGENESIS.  
RX MEDLINE=99263519; Pubmed=10329126;  
RA Kwak Y.T., Ivanov D., Guo J., Nee E., Gaynor R.B.;  
RT "Role of the human and murine cyclin T proteins in regulating HIV-1  
RT Tat-activation.";  
RT J. Mol. Biol. 288:57-69(1999).  
RN [5]  
RP FUNCTION.  
RX MEDLINE=99051315; Pubmed=9832504;  
RA Garber M.E., Wei P., Kewatamani V.N., Mayall T.P., Herrmann C.H.,  
RA Rice A.P., Littman D.R., Jones K.A.;  
RT "The interaction between HIV-1 Tat and human cyclin T1 requires zinc  
RT and a critical cysteine residue that is not conserved in the murine  
RT Cycl1 protein.";  
RT Genes Dev. 12:3512-3527(1998).  
RN [6]  
RP FUNCTION.  
RX MEDLINE=99292873; Pubmed=10364329;  
RA Bieniasz P.D., Grdina T.A., Bogerd H.P., Cullen B.R.;  
RT "Analysis of the effect of natural sequence variation in Tat and in  
RT cyclin T on the formation and RNA binding properties of Tat-cyclin T  
RT complexes.";  
RT J. Virol. 73:5777-5786(1999).  
RN [7]  
RP -1- FUNCTION: REGULATORY SUBUNIT OF THE CYCLIN-DEPENDENT KINASE PAIR  
RP (CDK9/CYCLIN T) COMPLEX, ALSO CALLED POSITIVE TRANSCRIPTION  
RP ELONGATION FACTOR B (P-TEFB), WHICH IS PROPOSED TO FACILITATE THE  
RP TRANSITION FROM ABORTIVE TO PRODUCTION ELONGATION BY  
RP PHOSPHORYLATING THE CTD (CARBOXY-TERMINAL DOMAIN) OF THE LARGE  
RP SUBUNIT OF RNA POLYMERASE II (RNAP II). BINDS ALSO TO THE  
RP TRANSCRIPTIONAL DOMAIN OF THE HIV-1 NUCLEAR TRANSCRIPTIONAL  
RP ACTIVATOR, TAT. IT IS PROBABLY THE CIS-ACTING TRANSCRIPTIONAL  
RP RESPONSE ELEMENT (TAR) RNA-BINDING COFACTOR FOR TAT. ALSO SERVES  
RP AS AN ESSENTIAL COFACTOR FOR HIV 2-TAT AND SIMIAN IMMUNODEFICIENCY  
RP VIRUS (AFRICAN MANDRILL) TAT FUNCTION.  
RP -1- SUBUNIT: ASSOCIATES WITH CDK9 TO FORM P-TEFB. CYCLIN T1 IS THE  
RP PREDOMINANT CYCLIN ASSOCIATED WITH CDK9.  
RP -1- SUBCELLULAR LOCATION: NUCLEAR.  
RP -1- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED.  
RP -1- MISCELLANEOUS: INTERACTION BETWEEN TAT AND CYCLIN T1 REQUIRES  
RP ZINC.  
RP -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN C SUBFAMILY.  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF045161; AAC39638.1; -  
DR EMBL; AF048730; AAC39664.1; -



DR MIM: 602506; -  
 DR InterPro: IPR000553; CYCLIN.  
 DR SMART: SM00385; CYCLIN: 1.  
 DR PROSITE: PS00292; CYCLINS; FALSE\_NEG.  
 KM Cyclin: Cell cycle; Cell division; Coiled coil;  
 KW Transcription regulation; Nuclear protein.  
 FT DOMAIN 253 270  
 FT DOMAIN 254 272 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 384 425 TAT: TAR RECOGNITION MOTIF (TRM).  
 FT DOMAIN 517 526 COILED COIL (POTENTIAL).  
 FT DOMAIN 560 570 POLY-HIS.  
 FT DOMAIN 717 725 POLY-SER.  
 FT MUTAGEN 261 261 C->Y: LOSS OF HIV-1 TAT TRANSACTIVATION.  
 FT CONFLICT 77 77 Q -> R (IN REF. 2).  
 SQ SEQUENCE 726 AA; 80664 MW; 4637FEB2DEDEF13 CRC64;

Query Match 69.7%; Score 53; DB 1; Length 726;  
 Best Local Similarity 80.0%; Pred. No. 33;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13  
 DB 717 PPPPLPPPK 726

RESULT 13  
 CCT1\_HORSE STANDARD; PRT; 727 AA.  
 ID C9XT26;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CYCLIN T1 (CYCLIN T) (CYCT1).  
 GN CCNT1.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=99303751; PubMed=10373508;  
 RA Bieniasz P.D., Grdina T.A., Bogerd H.P., Cullen B.R.;  
 RT "Highly divergent lentiviral tat proteins activate viral gene  
 expression by a common mechanism".  
 RL Mol. Cell. Biol. 19:4592-4599(1999).  
 CC -1- FUNCTION: REGULATORY SUBUNIT OF THE CYCLIN-DEPENDENT KINASE PAIR  
 (CDK9/CYCLIN T) COMPLEX, ALSO CALLED POSITIVE TRANSCRIPT  
 ELONGATION FACTOR B (P-TEFB), WHICH IS PROPOSED TO FACILITATE THE  
 TRANSITION FROM ABORTIVE TO PRODUCTION ELONGATION BY  
 PHOSPHORYLATING THE CTD (CARBOXY-TERMINAL DOMAIN) OF THE LARGE  
 SUBUNIT OF RNA POLYMERASE II (RNAP II) (BY SIMILARITY). BINDS ALSO  
 TO THE TRANSACTIVATION DOMAIN OF THE EQUINE INFECTIOUS ANEMIA  
 VIRUS (EIAV) NUCLEAR TRANSCRIPTIONAL ACTIVATOR. TAT. IT IS  
 PROBABLY THE CIS-ACTING TRANSACTIVATION RESPONSE ELEMENT (TAR)  
 RNA-BINDING COFACTOR FOR TAT. DOES NOT BIND TO THE TRANSACTIVATION  
 DOMAIN OF THE HIV-1 NUCLEAR TRANSCRIPTIONAL ACTIVATOR.  
 CC -1- SUBUNIT: ASSOCIATES WITH CDK9 TO FORM P-TEFB. CYCLIN T1 IS THE  
 PREDOMINANT CYCLIN ASSOCIATED WITH CDK9.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN C SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AF137509; AAD38518.1; -  
 DR InterPro: IPR000553; Cyclin.

DR SMART: SM00385; CYCLIN; 1.  
 DR PROSITE: PS00292; CYCLINS; FALSE\_NEG.  
 KM Cyclin: Cell cycle; Cell division; Coiled coil;  
 KW Transcription regulation; Nuclear protein.  
 FT DOMAIN 253 270  
 FT DOMAIN 254 272 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 386 427 TAT: TAR RECOGNITION MOTIF (TRM) (BY  
 SIMILARITY).  
 FT DOMAIN 519 528 COILED COIL (POTENTIAL).  
 FT DOMAIN 562 573 POLY-HIS.  
 FT DOMAIN 718 726 POLY-SER.  
 FT DOMAIN 726 726 POLY-PRO.  
 SQ SEQUENCE 727 AA; 81013 MW; BFC2A398DB35BCE CRC64;

Query Match 69.7%; Score 53; DB 1; Length 727;  
 Best Local Similarity 80.0%; Pred. No. 33;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13  
 DB 718 PPPPLPPPK 727

RESULT 14  
 FASL\_HUMAN STANDARD; PRT; 281 AA.  
 ID F48023;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE FAS ANTIGEN LIGAND (APOPTOSIS ANTIGEN LIGAND) (APTL).  
 GN TNFRSF6 OR APTL1G1 OR FASL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RC TISSUE=Thymus;  
 RX MEDLINE=95105731; PubMed=7528780;  
 RA Alderson M.;  
 RT "Fas ligand mediates activation-induced cell death in human T  
 lymphocytes".  
 RL J. Exp. Med. 181:71-77(1995).  
 CC -1- FUNCTION: CYTOKINE THAT BINDS TO FAS ANTIGEN, A RECEPTOR THAT  
 TRANSDUCES THE APOPTOTIC SIGNAL INTO CELLS. MAY BE INVOLVED IN

```

CC CYTOTOXIC T CELL MEDIATED APOPTOSIS AND IN T CELL DEVELOPMENT.
CC FAS-ANTIGEN MEDIATED APOPTOSIS MAY HAVE A ROLE IN THE INDUCTION OF
CC PERIPHERAL TOLERANCE, IN THE ANTIGEN-STIMULATED SUICIDE OF MATURE
CC T CELLS, OR BOTH.
CC -1- SUBUNIT: HOMOTRIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MAY BE RELEASED
CC INTO THE EXTRACELLULAR FLUID, PROBABLY BY CLEAVAGE FROM THE CELL
CC SURFACE.
CC -1- SIMILARITY: BELONGS TO THE TUMOR NECROSIS FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X89102; CAA61474.1; -
DR EMBL; U08137; AAC50071.1; -
DR EMBL; U11821; AAC50124.1; -
DR EMBL; D38122; BAA07320.1; -
DR EMBL; Z96050; CAB09424.1; -
DR EMBL; AB013303; BAA32542.1; -
DR HSSP; P01375; 2TUN.
DR MIM; 134638; -
DR InterPro; IPR003263; TNF_5.
DR InterPro; IPR000478; TNF_family.
DR Pfam; PF00229; TNF; 1.
DR ProDom; PD008600; TNF_5; 1.
DR SMART; SM00207; TNF_1.
DR PROSITE; PS00251; TNF_1; 1.
DR CycloLine; Transmembrane; Glycoprotein; Signal-anchor; Apoptosis.
KW CycloLine; Transmembrane; Glycoprotein; Signal-anchor; Apoptosis.
FT DOMAIN 1 80 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 81 102 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 103 281 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 4 70 PRO-RICH.
FT DISULFID 45 65 POLY-PRO.
FT CARBOHYD 202 233 BY SIMILARITY.
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 281 AA; 31485 MW; ABA6EB358246E9BB CRC64;

Query Match 68.4%; Score 52; DB 1; Length 281;
Best Local Similarity 88.9%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12
   |||||
Db 50 PPPPLPPPP 58

RESULT 15
SMN_MOUSE STANDARD: PRT: 288 AA.
AC P97801; 009092;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SURVIVAL MOTOR NEURON PROTEIN.
GN SMN1 OR SMN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97224505; PubMed=9070939;
RA Viollet L., Bertrand S., Brunialti A.L.B., Lefebvre S., Burlet P.,
RA Clermont O., Cruaud C., Guenet J.-L., Munnich A., Melki J.;

```

```

RT "cDNA isolation, expression, and chromosomal localization of the
RT mouse survival motor neuron gene (Smn)".;
RL Genomics 40:185-188(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97264340; PubMed=9110173;
RA Diodato C.J., Chen X.N., Noya D., Korenberg J.R., Nadeau J.H.,
RA Simard L.R.;
RT "Cloning, characterization, and copy number of the murine survival
RT motor neuron gene: homolog of the spinal muscular atrophy-determining
RT gene".;
RL Genome Res. 7:339-352(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Brain;
RX MEDLINE=97420785; PubMed=9275227;
RA Schrank B., Goetz R., Gunnersen J.M., Ure J.M., Toyka K.V.,
RA Smith A.G., Sendtner M.;
RT "Inactivation of the survival motor neuron gene, a candidate gene for
RT human spinal muscular atrophy, leads to massive cell death in early
RT mouse embryos.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:9920-9925(1997).
CC -1- FUNCTION: THE SMN COMPLEX PLAYS AN ESSENTIAL ROLE IN SPliceosomal
CC SNNRP ASSEMBLY IN THE CYTOPLASM, AND IS REQUIRED FOR PRE-mRNA
CC SPLICING IN THE NUCLEUS (BY SIMILARITY).
CC -1- SUBUNIT: FORMS A STABLE HETEROMERIC COMPLEX WITH SURVIVAL OF MOTOR
CC NEURON PROTEIN INTERACTING PROTEIN 1 (SIP1/GEMIN2), GEMIN3 AND
CC GEMIN4. THIS COMPLEX IS THOUGHT TO PLAY AN ESSENTIAL ROLE IN
CC BIOGENESIS OF THE SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLES
CC (SNNRPs). INTERACTS WITH FIBRILLARIN AND WITH SEVERAL SPliceosomal
CC SNNRP CORE SM PROTEINS, INCLUDING E/B', SM D AND SM E PROTEINS (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR. LOCALIZED IN
CC SUBNUCLEAR STRUCTURES NEXT TO COILED BODIES, CALLED GEMs, WHICH
CC ARE HIGHLY ENRICHED IN SPliceosomal SNNRPs (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SMN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U63294; AAC53057.1; -
DR EMBL; U77714; AAC53144.1; -
DR EMBL; Y12835; CAA73356.1; -
DR MGD; MGI:109257; Smn.
DR InterPro; IPR002999; Tudor.
DR SMART; SM00333; TUDOR; 1.
KW mRNA processing; RNA-binding; Nuclear protein.
FT DOMAIN 190 196 POLY-PRO.
FT DOMAIN 212 222 POLY-PRO.
FT DOMAIN 239 243 POLY-PRO.
SQ SEQUENCE 288 AA; 31254 MW; 757B3074649F7458 CRC64;

Query Match 68.4%; Score 52; DB 1; Length 288;
Best Local Similarity 88.9%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12
   |||||
Db 214 PPPPLPPPP 222

```

Search completed: January 22, 2002, 16:30:29  
Job time: 339 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:30:02 ; Search time 40.9 Seconds  
(without alignments)  
46.492 Million cell updates/sec

Title: US-09-485-529-103  
Perfect score: 76  
Sequence: 1 LNAPPPPLPPAPQ 13

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mhc:\*  
8: SP-organelle:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-virus:\*  
13: SP-vertebrate:\*  
14: SP-unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76	100.0	623	10 Q9ST59	Q9ST59 tritlicum ae
2	59	77.6	180	5 Q9UB2	Q9UB2 leishmania
3	59	77.6	625	10 Q9MB96	Q9MB96 oryza sativ
4	57	75.0	282	4 Q9BSA1	Q9BSA1 homo sapien
5	57	75.0	407	2 Q9AAS9	Q9AAS9 caulobacter
6	57	75.0	862	5 Q44708	Q44708 drosophila
7	57	75.0	956	5 Q9GU21	Q9GU21 drosophila
8	57	75.0	1214	4 Q9BQ04	Q9BQ04 homo sapien
9	56	73.7	108	5 Q9VNZ6	Q9VNZ6 drosophila
10	56	73.7	391	2 Q9RFX5	Q9RFX5 bruceella su
11	56	73.7	599	10 P93797	P93797 volvox cart
12	56	73.7	630	10 Q9ST48	Q9ST48 zea mays (m
13	56	73.7	804	4 Q9UPS6	Q9UPS6 homo sapien
14	56	73.7	1943	11 Q99P01	Q99P01 mus musculu
15	55	72.4	267	5 Q18233	Q18233 caenorhabdi
16	55	72.4	369	11 Q9QYD1	Q9QYD1 mus musculu
17	55	72.4	372	11 Q99MX7	Q99MX7 mus musculu
18	55	72.4	579	10 Q9LGG8	Q9LGG8 oryza sativ
19	55	72.4	770	5 Q44014	Q44014 leishmania

20	55	72.4	1175	6 Q9TW66	Q9TW66 oryctolagus
21	54	71.1	67	11 Q9Z118	Q9Z118 rattus norv
22	54	71.1	138	10 Q9SSK0	Q9SSK0 arabidopsis
23	54	71.1	263	5 Q9VWK3	Q9VWK3 drosophila
24	54	71.1	302	5 Q9VND7	Q9VND7 drosophila
25	54	71.1	318	5 Q9VHT1	Q9VHT1 drosophila
26	54	71.1	375	11 Q99J35	Q99J35 mus musculu
27	54	71.1	451	11 Q62775	Q62775 rattus norv
28	54	71.1	485	11 Q92068	Q92068 rattus norv
29	54	71.1	780	10 Q9LKT8	Q9LKT8 arabidopsis
30	54	71.1	839	5 Q9Y103	Q9Y103 drosophila
31	54	71.1	839	5 Q9NHN7	Q9NHN7 drosophila
32	54	71.1	853	4 Q14333	Q14333 homo sapien
33	54	71.1	988	5 Q01864	Q01864 caenorhabdi
34	54	71.1	1043	5 Q9VDC6	Q9VDC6 drosophila
35	53	69.7	203	2 Q9RXC9	Q9RXC9 streptomyce
36	53	69.7	237	10 Q9FIP7	Q9FIP7 arabidopsis
37	53	69.7	325	10 Q9LKB7	Q9LKB7 oryza sativ
38	53	69.7	373	2 Q53253	Q53253 mycobacteri
39	53	69.7	385	5 Q9N5D3	Q9N5D3 caenorhabdi
40	53	69.7	409	10 Q9SBM1	Q9SBM1 volvox cart
41	53	69.7	420	13 Q9DEG1	Q9DEG1 gallus gall
42	53	69.7	507	10 Q9FUE2	Q9FUE2 oryza sativ
43	53	69.7	513	13 Q9DEG2	Q9DEG2 gallus gall
44	53	69.7	550	13 Q93263	Q93263 gallus gall
45	53	69.7	561	4 Q92952	Q92952 homo sapien

## ALIGNMENTS

RESULT	ID	Q9ST59	PRELIMINARY:	PRT:	623 AA.
AC	Q9ST59				
DT	01-MAY-2000 (TReMBLrel. 13, Created)				
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)				
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)				
DE	GIBBERELLIN RESPONSE MODULATOR.				
GN	RHT-D1A.				
OS	Triticum aestivum (Wheat).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;				
OC	Triticeae; Triticum.				
OX	NCBI_TaxID=4565;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE-99347734; PubMed-10421366;				
RA	Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,				
RA	Flintham J.E., Beales J., Fish L.U., Worland A.D., Pelica F.,				
RA	Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;				
RT	'Green revolution' genes encode mutant gibberellin response				
RT	modulators.";				
RL	Nature 400:256-261(1999).				
DR	EMBL: AJ242531; CAB51555.1; -				
SQ	SEQUENCE 623 AA; 65337 MW; F2BAC34996D0A84F CRC64;				
Query Match 100.0%; Score 76; DB 10; Length 623;					
Best local Similarity 100.0%; Pred. No. 0.013; Mismatches 0; Indels 0; Gaps 0;					
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1 LNAPPPPLPPAPQ 13				
DB	109 LNAPPPPLPPAPQ 121				
RESULT	2				
ID	Q9UB2	PRELIMINARY:	PRT:	180 AA.	
AC	Q9UB2				
DT	01-MAY-2000 (TReMBLrel. 13, Created)				
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)				

DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE HYPOTHEITICAL 20.3 KDA PROTEIN.  
 GN L2719.02.  
 OS Leishmania major.  
 CC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA Fuchs M., Gabel C., Mueller-Auer S., Schaefer M., Rieger M.,  
 Ravens A.C., Lawson D., Quail M., Rajandream M.A., Barrell B.G.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA MEDLINE=98146435; PubMed=9477341;  
 RX Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
 Smith D.F.;  
 RA "A physical map of the Leishmania major Friedlin genome.";  
 RL Genome Res. 8:135-145(1998).  
 DR EMBL: AL117324; CAB55537.1; -.  
 KW Hypothetical protein  
 SQ SEQUENCE 180 AA; 20276 MW; 1F7E989F71E3A8CD CRC64;

Query Match 77.6%; Score 59; DB 5; Length 180;  
 Best Local Similarity 90.9%; Pred. No. 0.69;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 LNAPPPPLPPA 11  
 DB 83 LNAPPPPTPPA 93

RESULT 3  
 Q9MB96 PRELIMINARY; PRT; 625 AA.  
 AC Q9MB96;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE OSGAI.  
 GN OSGAI.  
 OS Oryza sativa (Rice).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 CC Ehrhartoideae; Oryzae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, NIPPONBARE;  
 RX MEDLINE=20179680; PubMed=10713441;  
 RA Ogawa M., Kusano T., Katsumi M., Sano H.;  
 RT "Rice gibberellin-insensitive gene homolog, OSGAI, encodes a nuclear-  
 RT localized protein capable of gene activation at transcriptional  
 RT level.";  
 RL Gene 245:21-29(2000).  
 DR EMBL: AB030956; BAA90749.1; -.  
 SQ SEQUENCE 625 AA; 65406 MW; 034FF02719DA2E97 CRC64;

Query Match 77.6%; Score 59; DB 10; Length 625;  
 Best Local Similarity 83.3%; Pred. No. 2.1;  
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LNAPPPPLPPAP 12  
 DB 110 LNAPPLPPAP 121

RESULT 4  
 Q9BSA1 PRELIMINARY; PRT; 282 AA.

AC Q9BSA1;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE ACTIVATING TRANSCRIPTION FACTOR 5.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CHORIOCARCINOMA;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC005174; AAH05174.1; -.  
 SQ SEQUENCE 282 AA; 30690 MW; B6D4CBF631655B1 CRC64;

Query Match 75.0%; Score 57; DB 4; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
 DB 131 PPPPLPPAP 139

RESULT 5  
 Q9AA59 PRELIMINARY; PRT; 407 AA.  
 AC Q9AA59;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE OMPA FAMILY PROTEIN.  
 GN CC0747.  
 OS Caulobacter crescentus.  
 CC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 CC Caulobacter.  
 OX NCBI\_TaxID=69394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 Potočka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 Kolonay J.F., Smit J., Craven M.B., Khoult H., Shetty J., Berry K.,  
 Ueberback T., Tian K., Wolf A., Yamalhevyan J., Ermolaeva M., White O.,  
 Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL: AE005750; AAK22732.1; -.  
 DR TIGR: CC0747; -.  
 KW Complete proteome.  
 SQ SEQUENCE 407 AA; 42781 MW; 54610D5FA64CC3BF CRC64;

Query Match 75.0%; Score 57; DB 2; Length 407;  
 Best Local Similarity 100.0%; Pred. No. 2.6;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
 DB 260 PPPPLPPAP 268

RESULT 6  
 O44708 PRELIMINARY; PRT; 862 AA.  
 AC O44708; Q9VE64; Q9VE65;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last sequence update)

DT 01-JUN-2001 (TREMBLrel\_17, last annotation update)  
DE FRUITLESS PROTEIN, ISOFORMS MALE, FEMALE, 3 AND 4 (FRUITLESS CLASS I  
DE MALE ISOFORM) (FRUITLESS CLASS I FEMALE ISOFORM) (CG7690 PROTEIN).  
GN F80 OR CG7689 OR CG7689 OR CG7690.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Tracheata; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_taxonomy:2271;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS MALE AND FEMALE).  
RP STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RX Ananthites P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RX George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RX Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RX Brandon R.C., Rogers Y.-H.C., Blake R.G., Champe M., Pfeiffer B.D.,  
RX Wen K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RX April J.F., Agayuni A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
RX Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RX Beeson K.Y., Benos P.V., Berman B.P., Bhendari D., Bolshakov S.,  
RX Borokova D., Botchan M.R., Bouck Y., Brokstein P., Brothier P.,  
RX Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RX Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RX de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RX Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RX Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RX Fodde C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RX Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RX Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RX Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,  
RX Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RX Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RX Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RX Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RX Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RX Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RX Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclio J.M.,  
RX Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,  
RX Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RX Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RX Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RX Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RX Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RX Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RX Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RX Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RX Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RT Science 287:2185-2195(2000).  
CC -I- ALTERNATIVE PRODUCTS: 5 ISOFORMS; 3 (SHOWN HERE), 4, MALE, FEMALE  
CC AND 5 (AC P91618); ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC EMBL: AF039931; AAB96677.1; -;  
DR EMBL: AF003722; AAF55564.1; -;  
DR EMBL: AEO03722; AAF55565.1; ALT\_TERM.  
DR EMBL: U72492; AAB92662.1; -;  
DR FlyBase; FBgn0004652; fru.  
DR InterPro; IPR000210; BTB\_POZ.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR000823; znf-C2H2.  
DR Pfam; PF00651; BTB\_1.  
DR Pfam; PF00096; zf-C2H2; 1.  
DR SMART; SM00225; BTB; 1.

DR	PROSITE: PSS0097; BTB: 2.
DR	PROSITE: PSS0186; EGF-2: 1.
KM	Alternative splicing; DNA-binding; EGF-like domain; Glycoprotein;
KM	Zinc-finger.
FT	VASPLIC 1 101 MISSING (IN FEMALE ISOFORM AND ISOFORM 4).
PT	
PT	VASPLIC 324 336 MISSING (IN MALE ISOFORM AND FEMALE ISOFORM).
PT	
PT	VASPLIC 617 789
PT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	PCTMMRCRSCGEVTYNNMHVHSHTAQSRSMCPYPATYSRI DTLRSRLRVKHPRDLKLNSSI (IN ISOFORM MALE ISOFORM AND FEMALE ISOFORM) .
FT	
FT	VASPLIC 790 862 MISSING (IN MALE ISOFORM AND FEMALE ISOFORM).
SO	SEQUENCE 862 AA; 92593 MW; 94C229E991526701 CRC64;

Query Match	75.0%	Score 57	DB 5	Length 862
Best Local Similarity	75.0%	Pred. No. 4.9		
Matches	9	Conservative	1	Mismatches 2
				Indels 0
				Gaps 0
QY	1	LNAPPPPLPPAP	12	
	:		1	
Db	74	VRAPPPPLPPPP	85	

RESULT	7		
09G9U21			
ID	09G9U21	PRELIMINARY;	PRT; 956 AA.
AC	09G9U21;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	FRUITLESS TYPE-A.		
GN	FRU OR CG7688 OR CG7689 OR CG7690.		
OS	Drosophila melanogaster (fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_Taxid=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-CANTON S.		
RA	USU1-AOKI K., Ito H., Ui-Tet K., Takahashi K., Lukacsovich T.,		
RA	Awano W., Nakata H., Piao Z.F., Nilsson E.E., Tomida J., Yamamoto D.;		
RT	"Formation of the male-specific muscle in female Drosophila by ectopic		
RT	fruitless expression.";		
RL	Nat. Cell Biol. 0:0-0(2000).		
DR	EMBL: AF2201177; AAC825588.1; -		
DR	FlyBase; FBgn0004652; fru.		
DR	InterPro; IPR000210; BTB_POZ.		
DR	InterPro; IPR000822; znf-C2H2.		
DR	Pfam; PF00651; BTB; 1.		
DR	Pfam; PF00096; zf-C2H2; 1.		
DR	SMART; SM00355; znf.C2H2; 1.		
DR	PROSITE; PS50097; BTB; 1		
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.		
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1.		
DR	DNA-binding; Metal-binding; zinc-finger		
QW	SEQUENCE 956 AA; 103051 MW; 675260937C937FAF CRC64;		

Query Match	75.0%;	Score 57;	DB 5;	Length 956;
Best Local Similarity	75.0%;	Pred. No. 5.4;		
Matches	9;	Conservative	1;	Mismatches 2;
			Indels	0;
			Gaps	0;

```

OY 1 LNAPPPPPAP 12
   : |||||
DB 74 VRAPPPPPPPP 85

RESULT 8
O9B004 PRELIMINARY: PRT: 1214 AA.
ID O9B004
AC O9B004
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE KIA0298 PROTEIN.
GN KIA0298.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RT 7.
RT "Comparative sequencing of human chromosome 11p15 and mouse chromosome
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ400879; CAC35389.1;
SQ SEQUENCE 1214 AA; 134403 MW; A3973C0EA2D47C3A CRC64;

Query Match
Best Local Similarity 75.0%; Score 57; DB 4; Length 1214;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13
   : |||||
DB 465 PPPPLPPPPQ 474

RESULT 9
O9VN26 PRELIMINARY: PRT: 108 AA.
ID O9VN26
AC O9VN26
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE CG14571 PROTEIN.
GN CG14571.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN NCBI_TaxID=7227;
RP SEQUENCE FROM N.A.
RT STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed-10731132;
RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Barker E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouch J., Brokstein P., Brotlier P.,
RA Butulis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

```

```

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003595; AAF51766.1;
DR Flybase; FBgn0037118; CG14571.
SQ SEQUENCE 108 AA; 11926 MW; B67F87AC2B2D3A45 CRC64;

Query Match
Best Local Similarity 73.7%; Score 56; DB 5; Length 108;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 APPPLPPAP 12
   : |||||
DB 20 APPPLPPPPQ 29

RESULT 10
O9RPX5 PRELIMINARY: PRT: 391 AA.
ID O9RPX5
AC O9RPX5
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE VIRB10.
GN VIRB10.
OS Brucella suis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
RN NCBI_TaxID=29461;
RP SEQUENCE FROM N.A.
RT STRAIN-1330;
RX MEDLINE=99440171; PubMed-10510235;
RA O'Callaghan D., Cazevielle C., Allardet-Servent A., Bouy G.,
RA Poulongue V., Fritlos P., Kulakov Y., Ramuz M.;
RT "A homologue of the Agrobacterium tumefaciens VirB and Bordetella
RT pertussis Ptl Type IV secretion systems is essential for intracellular
RT survival of Brucella suis.";
RL Mol. Microbiol. 33:1210-1220(1999).
DR EMBL; AF14604; AAD56620.1;
SQ SEQUENCE 391 AA; 41453 MW; A6DC4AC9AFB8490D CRC64;

Query Match
Best Local Similarity 73.7%; Score 56; DB 2; Length 391;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 LNAPPPPPAP 12
   : |||||
DB 81 LPAPPPPPPPAP 92

RESULT 11
P93797 PRELIMINARY: PRT: 599 AA.
ID P93797

```



AC P93797;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE PHENOPHORIN-S PRECURSOR.  
 OS Volvox carterii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Volvocaceae; Volvox.  
 NCBI\_TaxID=3067;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-HK 10;  
 RC MEDLINE-97162277; PubMed-9009264;  
 RA Godl K., Hallmann A., Wenzl S., Sumper M.;  
 RT "Differential targeting of closely related ECM glycoproteins: the  
 RL pherophorin family from Volvox.";  
 EMBO J. 16:25-34(1997).  
 DR EMBL; Y07752; CAA69032.1;  
 DR Mendei; 14875; Volca;2392;14875.  
 KW Signal.  
 FT SIGNAL.  
 FT CHAIN 1 17 POTENTIAL.  
 FT SEQUENCE 599 AA; 63436 MW; 8BA1AD5A71F8082 CRC64;

Query Match  
 Best Local Similarity 73.7%; Score 56; DB 10; Length 599;  
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 NAPPPLPPAP 12  
 Db 216 NAPPSPLEPSP 226

RESULT 12  
 Q9ST48 PRELIMINARY; PRT; 630 AA.  
 AC Q9ST48;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE GIBBERELLIN RESPONSE MODULATOR (FRAGMENT).  
 GN D8.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Panicoideae; Andropogoneae; Zea.  
 NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99347734; PubMed-10421366;  
 RA Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,  
 RA Filitshak J.E., Beales J., Fish L.J., Worland A.J., Pelica F.,  
 RA Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;  
 RT "Green revolution" genes encode mutant gibberellin response  
 RT modulators.";  
 RL Nature 400:256-261(1999).  
 DR EMBL; AJ242530; CAB51557.1;  
 FT NON\_TER 630 630  
 FT SEQUENCE 630 AA; 66028 MW; 3D56851726C51042 CRC64;

Query Match  
 Best Local Similarity 73.7%; Score 56; DB 10; Length 630;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LNAPPPLPPA 11  
 Db 110 LNAPPPLPPA 120  
 RESULT 13  
 Q9UPPS6

ID Q9UPPS6 PRELIMINARY; PRT; 804 AA.  
 AC Q9UPPS6;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE KIAA1076 PROTEIN (FRAGMENT).  
 GN KIAA1076.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RC MEDLINE-99397452; PubMed-10470851;  
 RA Kikuno R., Nagase T., Ishikawa K., Hirosewa M., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIV.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 6:197-205(1999).  
 DR EMBL; AB028999; BA83028.1;  
 DR InterPro; IPR001214; SET.  
 DR InterPro; IPR003616; PostSET.  
 DR Pfam; PF00856; SET.1.  
 DR PROSITE; PS50280; SET.1.  
 DR SMART; SM00508; PostSET.1.  
 DR SMART; SM00317; SET.1.  
 FT NON\_TER 1  
 FT SEQUENCE 804 AA; 87997 MW; 13CAB0BA5420BE67 CRC64;

Query Match  
 Best Local Similarity 73.7%; Score 56; DB 4; Length 804;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 APPPLPPAP 12  
 Db 419 APPPLPPAP 428

RESULT 14  
 Q99PJ1 PRELIMINARY; PRT; 1943 AA.  
 AC Q99PJ1;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE PROTOCADHERIN.  
 GN AV.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20578757; PubMed-11138007;  
 RA Alagramam K.N., Murcia C.L., Kwon H.Y., Pawlowski K.S., Wright C.G.,  
 RA Woychik R.P.;  
 RT "The mouse ames waltzer hearing loss mutant is caused by mutation of  
 RT pcdh15, a novel protocadherin gene.";  
 RL Nat. Genet. 27:99-102(2001).  
 DR EMBL; AF281899; AAG53891.1;  
 FT SEQUENCE 1943 AA; 214816 MW; E3D7DB9F5F738652 CRC64;

Query Match  
 Best Local Similarity 73.7%; Score 56; DB 11; Length 1943;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 LNAPPPLPPAPQ 13  
 Db 1801 LSLPPPPPPAPR 1813

RESULT 15

018233

ID 018233 PRELIMINARY: PRT: 267 AA.

AC 018233:

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)

DE SIMILARITY WITH WILMS' TUMOR PROTEIN.

GN

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,

RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Geisel C., Steilyes L., Bradshaw H.,

RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RL Submitted (Mar-1996) to the EMBL/Genbank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Waterston R.;

RL Submitted (Nov-1996) to the EMBL/Genbank/DBJ databases.

DR EMBL; U53148; AAB37073.1; -;

SQ SEQUENCE 267 AA; 28592 MW; 039A28642CB6F175 CRC64;

Query Match 72.4%; Score 55; DB 5; Length 267;

Best Local Similarity 90.0%; Pred. No. 3.2;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13

DB 90 PPPPPPPAPQ 99

Search completed: January 22, 2002, 16:30:03  
Job time: 338 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:26:20 ; Search time 43.62 Seconds  
(without alignments)  
1057.947 Million cell updates/sec

Title: US-09-485-529-7  
Perfect score: 623  
Sequence: 1 MKREYQDAGSGGGGGGMS.....TIGWHTPLATISAWRIAGP 623

Scoring table:  
OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq\_1101.\*  
1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT.\*  
22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	623	100.0	623	20	AAV02540
2	425	68.2	425	20	AAV02539
3	132	21.2	630	20	AAV02541
4	71	11.4	138	20	AAV02544
5	29	4.7	256	20	AAV02538
6	23	3.7	123	20	AAV02543
7	19	3.0	259	18	AAW30794
8	19	3.0	262	18	AAW38194
9	19	3.0	277	18	AAW38193
10	19	3.0	282	18	AAW30795
11	19	3.0	517	21	AAW38577

12	19	3.0	518	21	AAW38576	Arabidopsis thaliana
13	19	3.0	531	21	AAW28575	Arabidopsis thaliana
14	19	3.0	532	18	AAW30792	Arabidopsis thaliana
15	19	3.0	532	22	AAE02545	A. thaliana trans
16	19	3.0	532	22	AAE01907	Arabidopsis thaliana
17	19	3.0	533	21	AAW38575	Arabidopsis thaliana
18	19	3.0	572	22	AAW31883	Arabidopsis thaliana
19	19	3.0	572	22	AAW31884	Amino acid sequenc
20	19	3.0	587	21	AAW28574	Arabidopsis thaliana
21	19	3.0	587	22	AAE02560	Arabidopsis thaliana
22	19	3.0	587	22	AAE01892	Arabidopsis thaliana
23	17	2.7	17	20	AAV02537	Peptide derived fr
24	14	2.2	100	20	AAV02542	Protein encoded by
25	14	2.2	221	19	AAW30793	Arabidopsis thaliana
26	13	2.1	1010	18	AAW34989	Arabidopsis thaliana
27	12	1.9	113	21	AAW32809	Teredinibacter end
28	12	1.9	833	21	AAW41835	EucaIyptus grandis
29	12	1.9	1136	20	AAV23330	Human OREF ORF1599
30	12	1.9	1136	20	AAV22655	Human tumour suppr
31	11	1.8	35	21	AAW11032	Human tumour thalia
32	11	1.8	216	21	AAW24090	Arabidopsis thaliana
33	11	1.8	216	21	AAW37177	Arabidopsis thaliana
34	11	1.8	231	22	AAW84954	Shrimp white spot
35	11	1.8	407	21	AAW60685	Arabidopsis thaliana
36	11	1.8	430	21	AAW17935	Arabidopsis thaliana
37	11	1.8	621	22	AAW84929	Shrimp white spot
38	11	1.8	797	20	AAW05850	Banana ripening fr
39	10	1.6	19	22	AAW17602	Peptide #4036 enco
40	10	1.6	19	22	AAW30122	Peptide #4159 enco
41	10	1.6	19	22	AAW05262	Peptide #3944 enco
42	10	1.6	25	22	AAW35702	Peptide #9739 enco
43	10	1.6	30	22	AAW21772	Peptide #8206 enco
44	10	1.6	30	22	AAW38096	Peptide #12133 enc
45	10	1.6	52	22	AAW32328	Peptide #6365 enco

#### ALIGNMENTS

RESULT 1	
AAV02540	standard; Protein; 623 AA.
XX	
AC	AAV02540;
XX	
DT	16-JUL-1999 (first entry)
XX	
DE	Protein encoded by wheat Rht clone 5a1 genomic sequence.
XX	
KW	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW	paclobutrazol.
XX	
OS	Triticum aestivum.
XX	
PN	WO9909174-A1.
XX	
PD	25-FEB-1999.
XX	
PF	07-AUG-1998; 98WO-GB02383.
XX	
PR	13-AUG-1997; 97GB-0017192.
XX	
PA	(PLAN-) PLANT BIOSCIENCE LTD.
XX	
PI	Harberd NP, Peng J, Richards DE;
XX	
DR	WPI; 1999-181040/15.
XX	
NR	N-PSDB; AAX36279.
PT	New Triticum aestivum polynucleotides - encode a polypeptide which
PT	provides inhibition of the growth of plants, which inhibition is
PT	antagonised by gibberellin, used to confer a dwarf phenotype



```

|||||
Db 241 svfemrlllaqpalakvlgvtravprivtvegeanhnsffldrfesllystmd 300
QY 499 SLEGGSSGGSPSVSSGAAPAACTDOVMSRYVLGRQICNVVAEGAERPRHHTLQ 558
Db 301 sleegssggpsvssgaapaaagldqvmesvylgrlcnvacegaerterhel19g 360
QY 559 WRNRLGNAGFEVHLGSMYKQASTLALFAGDGYKVEKEGCLTLGMHTRPLIATSAW 618
Db 361 wrnrlgnagfevhlgsnmykqastllalfagdgdykvekegcltlgmhttrpliatasw 420
QY 619 RLACP 623
Db 421 rlapg 425

```

## RESULT 3

AA02541  
ID AAY02541 standard; Protein; 630 AA.

AC AAY02541;

DT 16-JUL-1999 (first entry)

DE Protein encoded by maize 1a1 genomic clone sequence.

KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
KW paclobutrazol; maize.

OS Zea mays.

PN WO90909174-A1.

PD 25-FEB-1999.

PF 07-AUG-1998; 98WO-GB02383.

PR 13-AUG-1997; 97GB-0017192.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Harberd NP, Peng J, Richards DE;

DR WPI; 1999-181040/15.

DR N-PSDB; AAX36280.

PT New Triticum aestivum polynucleotides - encode a polypeptide which  
PT provides inhibition of the growth of plants, which inhibition is  
PT antagonised by gibberellin, used to confer a dwarf phenotype

PS Disclosure; Fig 9b; 88pp; English.

CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in Triticum aestivum, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence is encoded by  
CC the maize 1a1 genomic clone sequence.

SO Sequence 630 AA;

Query Match 21.2%; Score 132; DB 20; Length 630;  
Best Local Similarity 100.0%; Pred. No. 3.9e-105;

Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 293 PDSLLDAAFADLHAHFYESCPLKFAHPTANQALIEAFAGCRVHVDFGKOGMOP 352
Db 302 pdsllldaafadllhanfyescplykfanftanqalleafagcrvhwdfgkqmqwp 361
QY 353 ALLOALALRPGPPSRFLGVPQDETDALQOVGKMLAQFAHTIRVDFQYGLVAATL 412
Db 362 allqalalrpgppsrfltcvgppqdeida1qvgwklaqfahllrvdfqyglvaatl 421
QY 413 ADLEPMLQPEG 424
Db 422 adlepmlqpeg 433

```

## RESULT 4

AA02544  
ID AAY02544 standard; Protein; 138 AA.

AC AAY02544;

DT 16-JUL-1999 (first entry)

DE Protein encoded by the partial sequence of the wheat rht-10 allele.

KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
KW paclobutrazol.

OS Triticum aestivum.

PN WO90909174-A1.

PD 25-FEB-1999.

PF 07-AUG-1998; 98WO-GB02383.

PR 13-AUG-1997; 97GB-0017192.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Harberd NP, Peng J, Richards DE;

DR WPI; 1999-181040/15.

PT New Triticum aestivum polynucleotides - encode a polypeptide which  
PT provides inhibition of the growth of plants, which inhibition is  
PT antagonised by gibberellin, used to confer a dwarf phenotype

PS Disclosure; Fig 12b; 88pp; English.

CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in Triticum aestivum, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence is encoded by  
CC the partial sequence of the wheat rht-10 allele.

SO Sequence 138 AA;

Query Match 11.4%; Score 71; DB 20; Length 138;  
Best Local Similarity 100.0%; Pred. No. 3.3e-53;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	122	LNASTSSPTVTGSGGFNDLPSPVSSSSITATLPPIPSACATATPADIASDVSVPKRMKG	181
Db	68	LNASTSSPTVTGSGGFNDLPSPVSSSSITATLPPIPSACATATPADIASDVSVPKRMKG	127
Oy	182	GSSTSSSSSSSS 192	
Db	128	gsstssssss 138	
RESULT	5		
AA02538			
ID	AA02538	standard; Protein; 256 AA.	
XX			
AC	AA02538;		
XX			
DT	16-JUL-1999	(first entry)	
XX			
DE	Protein encoded by rice EST D39460 sequence.		
XX			
KW	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;		
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;		
XX	paclobutrazol; rice; expressed sequence tag; EST.		
XX			
OS	Oryza sativa.		
XX			
PN	MO9909174-A1.		
XX			
PD	25-FEB-1999.		
XX			
PF	07-AUG-1998;	98MO-GB02383.	
XX			
PR	13-AUG-1997;	97GB-0017192.	
XX			
PA	(PLAN-) PLANT BIOSCIENCE LTD.		
XX			
PI	Harberd NP, Peng J, Richards DE;		
XX			
DR	WPI: 1999-181040/15.		
XX			
DR	N-FSDB; AAX36277.		
XX			
PT	New Triticum Aestivum polynucleotides - encode a polypeptide which		
PT	provides inhibition of the growth of plants, which inhibition is		
XX	antagonised by gibberellin, used to confer a dwarf phenotype		
PS			
XX	Claim 12; Fig 6b; 88pp; English.		
XX			
CC	The specification describes polypeptides encoded by the Rht gene (and		
CC	its homologues) that, when expressed in Triticum aestivum, inhibit		
CC	growth of the plant. This growth inhibition is antagonised by		
CC	gibberellin. The products can be used to provide Rht expression in		
CC	plants, conferring a dwarf phenotype on a plant which is correctable		
CC	by treatment with gibberellin. In addition, the products can be		
CC	used to produce Rht mutant plants which are dwarfed compared with		
CC	wild-type, the dwarfing being gibberellin-insensitive. Taller plants		
CC	may be made by knocking out Rht or the relevant homologue gene in		
CC	the plant of interest. Plants may be made which are resistant to		
CC	compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,		
CC	e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds		
CC	dwarf but let crop plants grow tall. The present sequence is encoded by		
CC	rice expressed sequence tag (EST) AAD39460, which is homologous to the		
CC	wheat Rht gene.		
XX			
XQ	Sequence 256 AA;		

Query Match	4.7%;	Score 29;	DB 20;	Length 256;
Best Local Similarity	100.0%;	Pred. No. 7.9e-17;		
Matches 29;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	221	PVVVDVTQEAGIRLVHALLACAEAVQOEN	249
Db	227	PVVVDVTQEAGIRLVHALLACAEAVQGEN	255

RESULT	6
AAV02543	
ID	AAV02543 standard; Protein; 123 AA

DT 16-JUL-1999 (first entry)  
xx

DE Protein encoded by the partial sequence of the maize D8-2023 allele.

KW Rht gene; homologue; *Triticum aestivum*; wheat; growth inhibition;  
 KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
 KW paclobutrazol; maize.

OS Zea mays.

PN WO9909174-A1

PD 25-FEB-1999.

PF 07-AUG-1998; 98WO-GB02383.

PR 13-AUG-1997; 97GB-0017192.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Harberd NP, Peng J, Richards DE;

DR WPI; 1999-181040/15.  
DR N-DEDA; AY36397

[illegible]

PT provides inhibition of the growth of plants, which inhibition is antagonised by gibberellin, used to confer a dwarf phenotype

PS Disclosure; Flg 11d; 88pp; English.

The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in *Triticum Aestivum*, inhibit growth of the plant. This growth inhibition is antagonised by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds dwarf but let crop plants grow tall. The present sequence is encoded by the partial sequence of the maize D8-2023 allele.

**SQ**      **Sequence**      **123 AA;**

Query Match	3.7%	Score 23	DB 20	Length 123
Best Local Similarity	100.0%	Pred. No.	6.1e-12	
Matches 23	Conservative 0	Mismatches 0	Indels 0	Gaps 0

```
QY 51 SDMDADVAQKLEQLLEMANGMGVVG 73
    |||||
Db 35 sdmadvaqk1eq1emangmgvgvvg 57
```

RESULT	7
AAW30794	
ID	AAW30794 standard; Protein; 259 AA

AC	AAW30794;
XX	
DT	19-MAR-1998 (first entry)
XX	

DE	Arabidopsis thaliana gibberellin insensitivity gai-d2 gene product.
XX	
KW	Gibberellin insensitivity; gai-d2; plant growth inhibition;
KW	dwarf phenotype; lodging resistance; increased yield;
KW	flowering regulation; bolting inhibition; spinach; lettuce;
KW	antibody; identification; probe; primer; antisense; sense;
KM	expression regulation; co-suppression; rice;
KM	Bakane disease resistance.
OS	
XX	Arabidopsis thaliana.
XX	
PN	WO9729123-A2.
PD	
XX	14-AUG-1997.
XX	
PF	12-FEB-1997; 97MO-GH00390.
PR	
XX	12-FEB-1996; 96GB-0002796.
XX	
PA	(INNE-) INNES CENT INNOVATIONS LTD JOHN.
PI	
DR	Carol P, Harberd NP, Peng J, Richards DE;
DR	WPI: 1997-415295/38.
N-PSDB:	AAT91939.
PT	Nucleic acid encoding gibberellin inhibitor GAI and related
PT	antisense sequences - used to create tall, or particularly, dwarf
PT	plants, especially crops such as maize, rice and wheat
PS	
XX	Claim 17; Fig 6d; 76pp; English.
XX	
CC	The present sequence is the Arabidopsis thaliana gibberellin
CC	insensitivity (gai-d2) gene product (GAI-d2), the expression of
CC	which inhibits plant growth. However the inhibition is antagonised
CC	by gibberellin (GA), while gai expression confers a dwarf phenotype
CC	that is insensitive to GA. Manipulating gai-d2 and GAI-d2
CC	expression can produce tall or dwarf plants, particularly the
CC	latter for increased resistance to lodging and increased yield. It
CC	may also allow regulation of flowering, i.e. plants remain in the
CC	vegetative state until treated with GA, useful to inhibit bolting
CC	in spinach and lettuce. GAI-d2 can be used to raise specific
CC	antibodies for identifying homologous proteins or genes in other
CC	species. gai-d2 fragments can also be used as probes or primers to
CC	identify and clone related sequences, or in the preparation of
CC	antisense or sense expression regulants (co-suppressing)
CC	sequences. Rice plants that express GAI-d2 may be resistant to Bakane
CC	disease. Manipulation of gai-d2 and GAI-d2 makes it possible to
CC	tailor the degree of dwarfism and GA sensitivity to particular crops
CC	or situations.
XX	
XX	
SQ	Sequence 259 AA;
OY	
Query Match	3.0%; Score 19; DB 18; Length 259;
Best Local Similarity	100.0%; Pred. No. 3.2e-08;
Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	
314 CPYLKPAHTNOAILLEAF 332	
229 cpylktaftangalleaf 247	
RESULT 8	
ID AAM38194	
AAM38194 standard; Protein: 262 AA.	
AC AAM38194;	
DT 11-MAY-1998 (first entry)	
DE Arabidopsis SCARECROW SRPaz protein.	
XX	

KM	SCARECROW; SCR; SRPa2; transgenic plant; root; gravitropism;
KW	crop improvement.
XX	
OS	Arabidopsis thaliana.
PN	WO9741152-A1.
PD	
PP	06-NOV-1997.
XX	
PF	25-APR-1997; 97WO-US07022.
XX	
PR	24-APR-1997; 97US-0842445.
PR	26-APR-1996; 96US-0638617.
XX	
PA	(UYNX ) UNIV NEW YORK STATE.
XX	
PI	Benfey PN, Dilaurenzio L, Helariutta Y, Malamy JE;
PI	Pysh L, Wysocka-Diller J;
XX	
DR	WPI; 1997-549683/50.
PT	DNA encoding Arabidopsis SCARECROW protein - useful to modify plant
PT	cell division and therefore alter root development, or alter plant
PT	stem or hypocotyl gravitropism
XX	
XX	Claim 2; Page 114-115; 221pp; English.
XX	
CC	This is the amino acid sequence of Arabidopsis SRPa2, a protein
CC	that shows homology to Arabidopsis SCARECROW (SCR) protein (see
CC	AAW38178). SCR is a member of a novel protein family and contains a
CC	number of functional domains similar to those found in
CC	transcription factors. SCR is expressed specifically in embryo
CC	root progenitor tissue and in certain root and stem tissues. It
CC	regulates a specific asymmetric division, and controls gravitropic
CC	response in aerial structures and root formation. SCR proteins
CC	(see AAW38178-201) from dicots and monocots, including SCR proteins
CC	lacking 1-4 of MOTIFS I-VI, and polypeptides corresponding to MOTIF
CC	I, II, III, IV, V or VI of SCR are claimed. Transgenic plants can
CC	be engineered to overexpress the SCR protein, so that cell division
CC	is increased in roots, resulting in thicker root development, while
CC	a plant with an altered stem or hypocotyl gravitropism is less
CC	susceptible to lodging. Plants that contain an antisense molecule
CC	that suppresses the expression of endogenous SCR gene product have
CC	thinner root development (all claimed).
XX	
SQ	Sequence 262 AA;
XX	
Query Match	3.0%; Score 19; DB 18; Length 262;
Best Local Similarity	100.0%; Pred. No. 3.3e-08;
Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	314 CPYLKFAHFTANQATLEAF 332 
DB	123 cpylkfahftanqalleaf 141
RESULT 9	
ID	AAW38193
AAW38193	standard; Protein: 277 AA.
XX	
AC	AAW38193;
XX	
DT	11-MAY-1998 (first entry)
XX	
DE	Arabidopsis SCARECROW SRPa8 protein.
XX	
KW	SCARECROW; SCR; SRPa8; transgenic plant; root; gravitropism;
KW	crop improvement.
XX	
OS	Arabidopsis thaliana.
RH	Key Location/Qualifiers

```

FT Misc-difference 269
FT /note= "unidentified residue"
XX
XX MO9741152-A1.
XX
XX 06-NOV-1997.
XX
XX 25-APR-1997; 97WO-US07022.
XX
XX 24-APR-1997; 97US-0842445.
XX 26-APR-1996; 96US-0638617.
XX
XX (UNYV ) UNIV NEW YORK STATE.
XX
XX Benfey PN, Dilaurenzio L, Helariutta Y, Malamy JE;
XX Pysch L, Mysiocka-Diller J;
XX WPI; 1997-549683/50.
XX
XX DNA encoding Arabidopsis SCARECROW protein - useful to modify plant
XX cell division and therefore alter root development, or alter plant
XX stem or hypocotyl gravitropism
XX
XX Claim 2; Page 113-114; 221pp; English.
XX
XX This is the amino acid sequence of Arabidopsis SRP8, a protein
XX that shows homology to Arabidopsis SCARECROW (SCR) protein (see
XX AA0818178). SCR is a member of a novel protein family and contains a
XX number of functional domains similar to those found in
XX transcription factors. SCR is expressed specifically in embryo
XX root progenitor tissue and in certain root and stem tissues. It
XX regulates a specific asymmetric division, and controls gravitropic
XX response in aerial structures and root formation. SCR proteins
XX (see AA0818178-201) from dicots and monocots, including SCR proteins
XX lacking 1-4 of MOTIFS I-VI, and polypeptides corresponding to MOTIF
XX I, II, III, IV, V or VI of SCR are claimed. Transgenic plants can
XX be engineered to overexpress the SCR protein, so that cell division
XX is increased in roots, resulting in thicker root development, while
XX a plant with an altered stem or hypocotyl gravitropism is less
XX susceptible to lodging. Plants that contain an antisense molecule
XX that suppresses the expression of endogenous SCR gene product have
XX thinner root development (all claimed).
XX
XX Sequence 277 AA;
XX
XX
XX Query Match 3.0%; Score 19; DB 18; Length 277;
XX Best Local Similarity 100.0%; Pred. No. 3,4e-08;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 314 CPYLKFAHFTANQAILLEAF 332
XX ||||||||||||||||
DB 138 CPYLKfahftanqalleaf 156
XX
XX
XX RESULT 10
XX AA030795
XX ID AA030795 standard; Protein: 282 AA.
XX
XX AA030795;
XX
XX 19-MAR-1998 (first entry)
XX
XX Arabidopsis thaliana gibberellin insensitivity gai-d3 gene product.
XX
XX Gibberellin insensitivity; gai-d3; plant growth inhibition;
XX dwarf phenotype; lodging resistance; increased yield;
XX flowering regulation; bolting inhibition; spinach; lettuce;
XX antibody; identification; probe; primer; antisense; sense;
XX expression regulation; co-suppression; rice;
XX Bakane disease resistance.
XX Arabidopsis thaliana.
XX OS

```

```

XX
XX MO9729123-A2.
XX
XX 14-AUG-1997.
XX
XX 12-FEB-1997; 97WO-GB00390.
XX
XX 12-FEB-1996; 96GB-0002796.
XX
XX (INNE-) INNES CENT INNOVATIONS LTD JOHN.
XX
XX Carol P, Harberd NP, Peng J, Richards DE;
XX WPI; 1997-415295/38.
XX DR N-PSDB; AAT91940.
XX
XX Nucleic acid encoding gibberellin inhibitor GAI and related
XX antisense sequences - used to create tall, or particularly, dwarf
XX plants, especially crops such as maize, rice and wheat
XX
XX Claim 17; Fig 6f; 76pp; English.
XX
XX The present sequence is the Arabidopsis thaliana gibberellin
XX insensitivity (gai-d3) gene product (GAI-d3), the expression of
XX which inhibits plant growth. However the inhibition is antagonised
XX by gibberellin (GA), while gai expression confers a dwarf phenotype
XX that is insensitive to GA. Manipulating gai-d3 and GAI-d3
XX expression can produce tall or dwarf plants, particularly the
XX latter for increased resistance to lodging and increased yield. It
XX may also allow regulation of flowering, i.e. plants remain in the
XX vegetative state until treated with GA, useful to inhibit bolting
XX in spinach and lettuce. GAI-d3 can be used to raise specific
XX antibodies for identifying homologous proteins or genes in other
XX species. gai-d3 fragments can also be used as probes or primers to
XX identify and clone related sequences, or in the preparation of
XX antisense or sense expression regulating (co-suppressing)
XX sequences. Rice plants that express GAI-d3 may be resistant to Bakane
XX disease. Manipulation of gai-d3 and GAI-d3 makes it possible to
XX tailor the degree of dwarfism and GA sensitivity to particular crops
XX or situations.
XX
XX Sequence 282 AA;
XX
XX
XX Query Match 3.0%; Score 19; DB 18; Length 282;
XX Best Local Similarity 100.0%; Pred. No. 3,5e-08;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 314 CPYLKFAHFTANQAILLEAF 332
XX ||||||||||||||||
DB 229 CPYLKfahftanqalleaf 247
XX
XX
XX RESULT 11
XX AAG38577
XX ID AAG38577 standard; Protein: 517 AA.
XX
XX AAG38577;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 47612.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX OS
XX EPI033405-A2.
XX PM
XX 06-SEP-2000.
XX PD
XX

```



PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145813.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-015138.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.

PR 28-SEP-1999; 99US-0166458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158028.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 3.0%; Score 19; DB 21; Length 517;  
Best Local Similarity 100.0%; Pred. No. 5,9e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALILEAF 332  
|||||  
DB 231 CPYLKFAHFTANGALILEAF 249

## RESULT 12

ID AAG38576 standard; Protein; 518 AA.

AC AAG38576;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 47611.

KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PF 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 12-AUG-1999; 99US-0148319.  
PR 13-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0158294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 3.0%; Score 19; DB 21; Length 518;

Best Local Similarity 100.0%; Pred. No. 5.9e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 0;

OY 314 CPYLKFAHFTANQALIEAF 332  
|||||  
DB 232 CPYLKFAHFTANQALIEAF 250

RESULT 13  
AAB28575  
ID AAB28575 standard; Protein; 531 AA.  
XX  
AC AAB28575;  
XX  
DT 09-FEB-2001 (first entry)  
XX  
DE Arabidopsis SCL2.  
XX  
KW Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;  
KW transgenic plant; cell division; molecular marker; herbicide resistance;  
KW salt resistance; pathogen resistance; insect resistance.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO200053723-A2.  
XX  
PD 14-SEP-2000.  
XX  
PF 07-MAR-2000; 2000MO-US05875.  
XX  
PR 10-MAR-1999; 99US-0265585.  
XX  
PA (UYNY ) UNIV NEW YORK STATE.  
XX  
PI Bentley PN, Di Lorenzo J, Wysocka-Diller J, Malamy JE, Pysk L;  
PI Helariutta Y, Bruce W, Lim J;  
XX  
DR WPI; 2000-594315/56.  
XX  
PT Scarecrow gene useful for producing transgenic plants expressing genes  
PT whose product increases starch, lignin or cellulose biosynthesis and  
PT confers herbicide, pathogen or insect resistance -

```

XX Claim 14; Fig 13; 200pp; English.
PS
XX The present sequence is given in a specification relating to the
CC structure and function of a regulatory gene, SCARECROW (SCR).
CC SCARECROW-like (SCL) genes encoding proteins containing an amino acid
CC sequence similar to the sequence of MOR1F III (VHID) of Arabidopsis SCR
CC protein are disclosed. SCR and SCL genes are useful for producing
CC transgenic plants whose cell division is modified and root and/or stem
CC development and gravitropism of stem or hypocotyl is altered. Cell
CC division is increased or decreased in roots resulting in thicker or
CC thinner root development. The transgenic plants are useful for expressing
CC a gene of interest encoding a gene product that confers herbicide, salt,
CC pathogen or insect resistance in root or embryos and genes encoding
CC starch, lignin or cellulose biosynthesis in shoots. The SCR gene also
CC confers less susceptibility to lodging in the transgenic plants than a
CC wild-type plant. SCR gene sequences are also useful as molecular markers
CC for a quantitative trait e.g. root or gravitropism trait in molecular
CC breeding of crop plants.
XX
SQ Sequence 531 AA;

Query Match 3.0%; Score 19; DB 21; Length 531;
Best Local Similarity 100.0%; Pred. No. 6.1e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALTEAF 332
DB 245 CPYLKfahftanqalleaf 263

RESULT 14
AAM30792
ID AAM30792 standard; Protein; 532 AA.
XX
AC AAM30792;
XX
DT 19-MAR-1998 (first entry)
XX
DE Arabidopsis thaliana gibberellin insensitivity gai gene product.
XX
KW Gibberellin insensitivity; gai; plant growth inhibition;
KW dwarf phenotype; lodging resistance; increased yield;
KW flowering regulation; bolting inhibition; spinach; lettuce;
KW antibody; identification; probe; primer; antisense; sense;
KW expression regulation; co-suppression; rice;
KW Bakane disease resistance.
XX
XX Arabidopsis thaliana.
OS
XX
XX MO9729123-A2.
XX
XX 14-AUG-1997.
XX
XX 12-FEB-1997; 97WO-GB00390.
XX
XX 12-FEB-1996; 96GB-0002796.
XX
XX (INNE-) INNES CENT INNOVATIONS LTD JOHN.
XX
XX Carol P, Harberd NP, Peng J, Richards DE;
XX
XX WPI: 1997-415295/38.
XX
XX N-PSDB: AAT91937.
XX
XX Nucleic acid encoding gibberellin inhibitor GAI and related
XX antisense sequences - used to create tall, or particularly, dwarf
XX plants, especially crops such as maize, rice and wheat
XX
XX Claim 1; Fig 4; 76pp; English.
XX
XX The present sequence is the Arabidopsis thaliana gibberellin
CC

```

```

CC insensitivity (gai) gene product (GAI), the expression of which
CC inhibits plant growth. However the inhibition is antagonised by
CC gibberellin (GA), while gai expression confers a dwarf phenotype
CC that is insensitive to GA. Manipulating gai and GAI expression can
CC produce tall or dwarf plants, particularly the latter for increased
CC resistance to lodging and increased yield. It may also allow
CC regulation of flowering, i.e. plants remain in the vegetative state
CC until treated with GA, useful to inhibit bolting in spinach and
CC lettuce. GAI can be used to raise specific antibodies for
CC identifying homologous proteins or genes in other species. gai
CC fragments can also be used as probes or primers to identify and
CC clone related sequences, or in the preparation of antisense or
CC sense expression regulating (co-suppressing) sequences. Rice plants
CC that express GAI may be resistant to Bakane disease. Manipulation
CC of gai and GAI makes it possible to tailor the degree of dwarfism
CC and GA sensitivity to particular crops or situations.
XX
SQ Sequence 532 AA;

Query Match 3.0%; Score 19; DB 18; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.1e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALTEAF 332
DB 246 CPYLKfahftanqalleaf 264

RESULT 15
AAE02545
ID AAE02545 standard; Protein; 532 AA.
XX
AC AAE02545;
XX
DT 10-AUG-2001 (first entry)
XX
DE A. thaliana transcription factor G308.
XX
KW Plant transcription factor; phenotype; sugar sensing characteristic;
KW transgenic plant; plant yield; growth; germination; photosynthesis;
KW glyoxylate metabolism; respiration; pathogen response; wounding response;
KW cell cycle regulation; pigmentation; flowering; senescence; physiology;
XX
XX storage organ; metabolism.
XX
XX Arabidopsis thaliana.
OS
XX
XX Key Location/Qualifiers
XX Domain 270..274
XX /note="Conserved domain"

WO200135725-A1.
XX
XX 25-MAY-2001.
XX
XX 14-NOV-2000; 2000WO-US31414.
XX
XX 17-NOV-1999; 99US-0166228.
XX
XX 17-APR-2000; 2000US-0197899.
XX
XX 22-AUG-2000; 2000US-0227439.
XX
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
XX
XX (JIAN/) JIANG C.
XX
XX (HEAR/) HEARD J.
XX
XX (PINE/) PINEDA O.
XX
XX (PIUG/) PILGRIM M.
XX
XX (ADAM/) ADAM L.
XX
XX (RIEC/) RIECHMANN J L.
XX
XX (YUGG/) YU G.
XX
XX (SAMA/) SAMAHA R.
XX
XX JIANG C, Heard J, Pineda O, Pilgrim M, Adam L, Riechmann JL;
XX Yu G, Samaha R;
PI

```

XX WPI: 2001-335977/35.  
DR N-PSDB; AAD06646.

XX Nucleic acids encoding plant transcription factor polypeptides, useful  
PT for altering the sugar sensing characteristics of plants and increasing  
PT yield, e.g. corn, potato and cotton plants -

PS Claim 4; Page 74-76; 151pp; English.

XX  
CC The patent relates to polynucleotides encoding 35 plant transcription  
CC factors which may be used to modify phenotype associated with a plant's  
CC sugar sensing characteristics and increasing yield when their expression  
CC level is altered. Sugars are central regulatory molecules that control  
CC aspects of physiology, metabolism and development. Therefore the cDNAs  
CC and proteins of the invention are useful for modifying the growth and  
CC germination rates of plants, photosynthesis, glyoxylate metabolism,  
CC respiration, starch and sucrose synthesis and degradation, pathogen  
CC response, wounding response, cell cycle regulation, pigmentation,  
CC flowering and senescence of plants and for modifying sink-source  
CC relationships in seeds, tubers, roots, and other storage organs leading  
CC to an increase in yield. The transcription factor polynucleotides and  
CC polypeptides may be used to alter the structure and developmental  
CC characteristics of plants such as soybean, wheat, corn, potato, cotton,  
CC rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,  
CC blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower,  
CC coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,  
CC onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,  
CC tobacco, tomato, watermelon, rosaceous fruits and/or vegetable brassicas.  
CC The present sequence is an Arabidopsis thaliana transcription factor.

XX  
SQ Sequence 532 AA;

Query Match 3.0%; Score 19; DB 22; Length 532;

Best Local Similarity 100.0%; Pred. No. 6.1e-08;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPLYKFAHFTANQALILEAF 332

|||||

DB 246 CPLYKFAHFTANGALILEAF 264

Search completed: January 22, 2002, 16:31:44  
Job time: 324 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:29:20 ; Search time 26.82 seconds  
(without alignments)  
1769.454 Million cell updates/sec

Title: US-09-485-529-7

Perfect score: 623

Sequence: 1 MKREYDAGSGGGGGGSGMS.....TLGHTRPILATSAWRLGP 623

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: PIR.68:\*  
2: pir1:\*  
3: pir2:\*  
4: pir3:\*  
5: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	3.5	511	2	G96688
2	19	3.0	523	2	T51475
3	19	3.0	533	2	H86282
4	19	3.0	587	2	D84426
5	16	2.6	275	2	T04480
6	12	1.9	153	2	S67294
7	12	1.9	229	2	UC7219
8	12	1.9	390	2	E81408
9	12	1.9	446	2	T45525
10	12	1.9	475	2	S49886
11	11	1.8	164	2	T16168
12	11	1.8	216	2	B66794
13	11	1.8	238	2	T32889
14	11	1.8	429	2	E96685
15	11	1.8	575	2	T01552
16	11	1.8	632	2	T00679
17	11	1.8	693	2	S64904
18	11	1.8	816	2	S64439
19	11	1.8	1249	2	T62294
20	11	1.8	1772	2	A45532
21	10	1.6	205	2	T15450
22	10	1.6	210	2	T49785
23	10	1.6	230	2	T39598
24	10	1.6	238	2	T52505
25	10	1.6	240	2	D71272
26	10	1.6	242	2	D86343
27	10	1.6	246	2	S58497
28	10	1.6	262	2	T40941
29	10	1.6	265	2	D96667

30	10	1.6	284	2	S58650	hypothetical prote
31	10	1.6	306	2	T47832	hypothetical prote
32	10	1.6	349	2	T41394	hypothetical serin
33	10	1.6	406	2	T48103	mRNA binding prote
34	10	1.6	411	2	E96665	protein F22C12.16
35	10	1.6	412	2	T48405	myd-like protein -
36	10	1.6	414	2	D96838	unknown protein T2
37	10	1.6	421	2	S71331	L-ascorbate peroxi
38	10	1.6	430	2	T12282	L-ascorbate peroxi
39	10	1.6	443	2	T05540	hypothetical prote
40	10	1.6	461	2	T11819	glycerol-3-phospha
41	10	1.6	468	2	A55476	protein Kinase (EC
42	10	1.6	484	2	S65713	hypothetical prote
43	10	1.6	503	2	S63257	probable membrane
44	10	1.6	506	2	F86733	ABC transporter pe
45	10	1.6	568	2	I39411	AF-9 protein - hum

#### ALIGNMENTS

```

RESULT 1
G96688
hypothetical protein T27F4.10 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence-revision 02-Mar-2001 #text-change 31-Mar-2001
R:Theologos, A.; Ecker, J.R.; Palm, C.D.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L., Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MIDB:21016719
A:Accession: G96688
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-511 <STO>
A:Cross-references: GB:AE005173; NID:g10092507; PIDN:AG12907.1; GSPDB:GM00141
C:Genetics:
A:Gene: T27F4.10
A:Map position: 1

Query Match 3.5%; Score 22; DB 2; Length 511;
Best Local Similarity 100.0%; Pred. No. 4.7e-11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 309 HFE5CPYLFKFAHFTANQAILL 330
Db 225 HFESCPYLFKFAHFTANQAILL 246

RESULT 2
T51475
RGA-like protein - Arabidopsis thaliana
N:Alternate names: protein K3M16.60
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 18-Aug-2000 #sequence-revision 18-Aug-2000 #text-change 18-Aug-2000
C:Accession: T51475
R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.;
submitted to the Protein Sequence Database, August 2000
A:Reference number: Z25394
A:Accession: T51475
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-523 <SAT>
A:Cross-references: EMBL:AL391150
A:Experimental source: cultivar Columbia; BAC clone K3M16

```

C:Genetics:  
A:Map position: 5  
A:Note: K3M16\_60

Query Match 3.0%; Score 19; DB 2; Length 523;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 313 SCPYLKFAHFTANQALILEAF 331  
|||||  
Db 234 SCPYLKFAHFTANQALILEAF 252

RESULT 3  
H86282  
Protein F10B6.34 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: H86282

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hultzar, L.  
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: H86282

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-533 <STO>

A:Cross-references: GB:AE005172; NID:98778219; PIDN:AAF79228.1; GSPDB:GN00141

C:Genetics:

A:Gene: F10B6.34

A:Map position: 1

Query Match 3.0%; Score 19; DB 2; Length 533;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALILEAF 332  
|||||  
Db 247 CPYLKFAHFTANQALILEAF 265

RESULT 4  
D84426  
hypothetical protein At2g01570 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: D84426

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffit, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: D84426

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-587 <STO>

A:Cross-references: GB:AE002093; NID:93785986; PIDN:AAC67333.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g01570

A:Map position: 2

Query Match 3.0%; Score 19; DB 2; Length 587;

Best Local Similarity 100.0%; Pred. No. 2.2e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALILEAF 332  
|||||  
Db 299 CPYLKFAHFTANQALILEAF 317

RESULT 5  
T04480  
acyl-CoA oxidase homolog - barley (fragment)  
C:Species: Hordeum vulgare (barley)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 08-Oct-1999  
C:Accession: T04480

R:Grossi, M.; Gull, M.; Stanca, A.M.; Cattiveill, L.  
Plant Sci. 105, 71-80, 1995

A:Title: Characterization of two barley genes that respond rapidly to dehydration str

A:Reference number: 215371

A:Accession: T04480

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-275 <GRC>

A:Cross-references: EMBL:X84055; NID:9642243; PIDN:CA58874.1; PID:9642244

A:Experimental source: cv. Arda, leaf

C:Genetics:

A:Gene: cdt29

Query Match 2.6%; Score 16; DB 2; Length 275;  
Best Local Similarity 100.0%; Pred. No. 4.8e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 482 FLDRFTESLHYSTMF 497  
|||||  
Db 8 FLDRFTESLHYSTMF 23

RESULT 6  
S67294  
hypothetical protein YOR382w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein O6760

C:Species: Saccharomyces cerevisiae

C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 05-Nov-1999

C:Accession: S67294

R:Delius, H.; Hebling, U.; Hofmann, B.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67261

A:Accession: S67294

A:Molecule type: DNA

A:Residues: 1-153 <DEL>

A:Cross-references: EMBL:Z75290; NID:91420822; PID:e252449; PID:91420823; GSPDB:GN000

A:Experimental source: strain S288C

C:Genetics:

A:Gene: MIPS:YOR382w

A:Map position: 15R

Query Match 1.9%; Score 12; DB 2; Length 153;  
Best Local Similarity 100.0%; Pred. No. 0.009;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 SSTSSSSSSSS 194  
|||||  
Db 111 SSTSSSSSSSS 122

RESULT 7  
JC7219

nuclear protein SR-25 - mouse  
C:Species: Mus musculus (house mouse)

C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000

C:Accession: JC7219

R:Sasahara, K.; Yamaoka, T.; Moritani, M.; Tanaka, M.; Iwahana, H.; Yoshimoto, K.; Mi



Biochem. Biophys. Res. Commun. 269, 444-450, 2000  
A:Title: Molecular cloning and expression analysis of a putative nuclear protein, SR-25.  
A:Reference number: JG7219; MUID:20175222  
A:Accession: JG7219  
A:Molecule type: mRNA  
A:Residues: 1-229 <SAS>  
A:Cross-references: DDBJ:AB05383; NID:g7619895; PIDN:BAA94743.1; PID:g7619896  
A:Experimental source: MIN6 cell line  
C:Comment: This protein is a highly hydrophilic nuclear protein with a serine-arginine  
A:splicing factors.  
C:Keywords: nucleus; RNA processing

Query Match 1.9%; Score 12; DB 2; Length 229;  
Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 STSSSSSSSS 194  
|||||  
DB 69 STSSSSSSSS 80

RESULT 8  
E81408  
probable periplasmic protein Cj0606 [imported] - Campylobacter jejuni (strain NCTC 11168  
C:Species: Campylobacter jejuni  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 31-Mar-2000  
C:Accession: E81408  
R:Parhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barre  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A:Reference number: AB1250; MUID:20150912  
A:Accession: E81408  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-390 <PAR>  
A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB75242.1; PID:g696807  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: Cj0606

Query Match 1.9%; Score 12; DB 2; Length 390;  
Best Local Similarity 100.0%; Pred. No. 0.02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 STSSSSSSSS 194  
|||||  
DB 268 STSSSSSSSS 279

RESULT 9  
T45525  
WSC4 homolog [imported] - Yeast (Kluyveromyces marxianus var. lactis)  
C:Species: Kluyveromyces marxianus var. lactis; Candida spherica  
C>Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 31-Jan-2000  
C:Accession: T45525  
R: Bao, W.G.; Fukuhara, H.  
submitted to the EMBL Data Library, July 1999  
A:Description: The ubiquitin-encoding genes of Kluyveromyces lactis.  
A:Reference number: Z23000  
A:Accession: T45525  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-446 <BAO>  
A:Cross-references: EMBL:AJ243800; PIDN:CAB50897.1  
A:Experimental source: strain 2359/152  
C:Genetics:  
A:Gene: WSC4

Query Match 1.9%; Score 12; DB 2; Length 446;

Best Local Similarity 100.0%; Pred. No. 0.022;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 STSSSSSSSS 194  
|||||  
DB 184 STSSSSSSSS 195

RESULT 10  
S49886  
probable membrane protein Y1123w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein Y18277.06  
C:Species: Saccharomyces cerevisiae  
C>Date: 13-Jan-1995 #sequence\_revision 20-Feb-1995 #text\_change 11-Jan-2000  
C:Accession: S49886  
R: Hamlyn, N.; Churcher, C.  
submitted to the EMBL Data Library, November 1994  
A:Reference number: S49881  
A:Accession: S49886  
A:Molecule type: DNA  
A:Residues: 1-475 <HAM>  
A:Cross-references: GB:Z47047; EMBL:Z46833; NID:g603997; PID:g763223; MIPS:Y1123w  
C:Genetics:  
A:Gene: SGD:SIM1  
A:Cross-references: SGD:S0001385; MIPS:Y1123w  
A:Map position: 9L  
C:Superfamily: Saccharomyces NCA3 protein  
C:Keywords: transmembrane protein  
F:82-98/Domain: transmembrane #status predicted <TMM>

Query Match 1.9%; Score 12; DB 2; Length 475;  
Best Local Similarity 100.0%; Pred. No. 0.023;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 STSSSSSSSS 194  
|||||  
DB 182 STSSSSSSSS 193

RESULT 11  
T16168  
hypothetical protein F26F4.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T16168  
R: Fulton, L.  
submitted to the EMBL Data Library, March 1996  
A:Description: The sequence of C. elegans cosmid F26F4.  
A:Reference number: Z18471  
A:Accession: T16168  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-164 <FUL>  
A:Cross-references: EMBL:U12964; NID:g1213452; PID:g529204; PIDN:AAA91221.1; CESP:F26  
A:Experimental source: strain Bristol NZ  
C:Genetics:  
A:Gene: CESP:F26F4.6  
A:introns: 34/1

Query Match 1.8%; Score 11; DB 2; Length 164;  
Best Local Similarity 100.0%; Pred. No. 0.071;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 STSSSSSSSS 194  
|||||  
DB 106 STSSSSSSSS 116

RESULT 12  
B96794  
unknown protein F14G6.20 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: B96794  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpsiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Hulzar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzalla, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: B96794  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-216 <STO>  
 A:Cross-references: GB:AE005173; NID:96642677; PIDN:AAF20257.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F14G6.20  
 A:Map position: 1

Query Match 1.8%; Score 11; DB 2; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 0.089;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 STSSSSSSSS 194  
 Db 45 STSSSSSSSS 55

## RESULT 13

T32889  
 hypothetical protein C34B2.9 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T32889  
 R:Graves, T.; Suterer, C.; Hawkins, M.; Wilson, R.  
 A:Description: The sequence of C. elegans cosmid C34B2.  
 A:Reference number: 221241  
 A:Accession: T32889  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-238 <GBR>  
 A:Cross-references: EMBL:AF043693; PIDN:AAB97541.1; GSPDB:GN00019; CESP:C34B2.9  
 A:Experimental source: Strain Bristol N2; clone C34B2  
 C:Genetics:  
 A:Gene: CESP:C34B2.9  
 A:Map position: 1  
 A:Introns: 35/3; 91/2; 200/2

Query Match 1.8%; Score 11; DB 2; Length 238;  
 Best Local Similarity 100.0%; Pred. No. 0.096;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 STSSSSSSSS 194  
 Db 19 STSSSSSSSS 29

## RESULT 14

E96685  
 probable disease resistance protein F15E12.17 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: E96685  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpsiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Hulzar, L.

Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzalla Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: E96685  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-429 <STO>  
 A:Cross-references: GB:AE005173; NID:911038485; PIDN:AAG27764.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F15E12.17  
 A:Map position: 1

Query Match 1.8%; Score 11; DB 2; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 0.16;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 STSSSSSSSS 194  
 Db 4 STSSSSSSSS 14

## RESULT 15

T01552  
 hypothetical protein A\_TM018A10.2 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 24-Mar-1999  
 C:Accession: T01552  
 R:Dempsey, S.; Harper, M.  
 A:Description: The sequence of A. thaliana TM018A10.  
 A:Reference number: 214348  
 A:Accession: T01552  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-575 <DEM>  
 A:Cross-references: EMBL:AF013294; NID:g2252848; PID:g2252850  
 A:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 271/3  
 A:Note: A\_TM018A10.2

Query Match 1.8%; Score 11; DB 2; Length 575;  
 Best Local Similarity 100.0%; Pred. No. 0.2;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 STSSSSSSSS 194  
 Db 59 STSSSSSSSS 69

Search completed: January 22, 2002, 16:32:17  
 Job time: 177 sec

---





```

RESULT 2
YC18_HUMAN          STANDARD;          PRT;          864 AA.
AC  Q9ULX2;
DT  20-AUG-2001 (Rel. 40, Created)
DT  20-AUG-2001 (Rel. 40, Last sequence update)
DT  20-AUG-2001 (Rel. 40, Last annotation update)
DE  HYPOTHETICAL PROTEIN KIAA1218 (FRAGMENT).
GN  KIAA1218.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Brain;
RX  MEDLINE=20039619; PubMed=10574462;
RA  Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA  Ohara O.;
RT  "Prediction of the coding sequences of unidentified human genes. XV.
RT  The complete sequences of 100 new cDNA clones from brain which code
RT  for large proteins in vitro."
RL  DNA Res. 6:337-345(1999).
CC  -1- SIMILARITY: BELONGS TO THE ATAXIN 7 FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AB033044; BAA86532.1; -.
CC  DR  Hypothetical protein.
CC  KW  NON_TER
CC  FT  1
CC  FT  223 226 POLY-THR.
CC  FT  DOMAIN 652 664 POLY-SER.
CC  FT  DOMAIN 786 790 POLY-SER.
CC  SQ  SEQUENCE 864 AA; 92138 MW; D5CE8130E4BDA23 CRC64;

Query Match
Best Local Similarity 1.9%; Score 12; DB 1; Length 864;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 TSSSSSSSSSS 194
Db 652 TSSSSSSSSSS 663

RESULT 3
BDNF_CAVPO          STANDARD;          PRT;          255 AA.
AC  O70183;
DT  30-MAY-2000 (Rel. 39, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DT  30-MAY-2000 (Rel. 39, Last annotation update)
DE  BRAIN-DERIVED NEUROTROPHIC FACTOR PRECURSOR (BDNF).
GN  BDNF.
OS  Cavia porcellus (Guinea pig).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Hystriognathi; Cavidae; Cavia.
OX  NCBI_TaxID=10141;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=HARTLEY WHITE; TISSUE=Liver;
RA  Ioune M., Nakayama C., Noguchi H.;
RL  Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: PROMOTES THE SURVIVAL OF NEURONAL POPULATIONS THAT ARE
CC  ALL LOCATED EITHER IN THE CENTRAL NERVOUS SYSTEM OR DIRECTLY

```

```

CC  CONNECTED TO IT (BY SIMILARITY).
CC  -1- SUBCELLULAR LOCATION: SECRETED.
CC  -1- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AB012097; BAA25176.1; -.
CC  DR  InterPro: IPR002072; NGF.
CC  DR  Pfam: PF00243; NGF_1.
CC  DR  PRINTS: PD00268; NGF.
CC  DR  ProDom: PD002052; NGF; 1.
CC  DR  SMART: SM00140; NGF; 1.
CC  DR  PROSITE: PS00248; NGF_1; 1.
CC  DR  PROSITE: PS50270; NGF_2; 1.
CC  KW  Growth factor; Signal.
CC  FT  SIGNAL 1
CC  FT  PROPEP 19 136 POTENTIAL.
CC  FT  CHAIN 137 255 BY SIMILARITY.
CC  FT  DISULFID 149 216 BRAIN-DERIVED NEUROTROPHIC FACTOR.
CC  FT  DISULFID 194 245 BY SIMILARITY.
CC  FT  DISULFID 204 247 BY SIMILARITY.
CC  FT  CARBOHYD 129 129 N-LINKED (GLCNAC ...) (POTENTIAL).
CC  SQ  SEQUENCE 255 AA; 28308 MW; BA95BA3EBB8FA04 CRC64;

Query Match
Best Local Similarity 1.8%; Score 11; DB 1; Length 255;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 TSSSSSSSSSL 195
Db 57 TSSSSSSSSSL 67

RESULT 4
YG3A_YEAST          STANDARD;          PRT;          816 AA.
AC  P53278;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  01-OCT-1996 (Rel. 34, Last annotation update)
DE  HYPOTHETICAL 92.7 KDA PROTEIN IN ASN2-PH1 INTERGENIC REGION.
GN  YGR130C.
OS  Saccharomyces cerevisiae (Baker's yeast).
OC  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC  Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX  NCBI_TaxID=4932;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  van Dyck L., Skala J., de Wergifosse P., Purnelle B., Talia E.,
RA  Nawrocki A., del Bino S., Goffeau A.;
RL  Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; Z72915; CA97143.1; -.
CC  DR  HSSP: P56255; LPJR.
CC  DR  SGD: S0003362; YGR130C.
CC  KW  Hypothetical protein.
CC  FT  DOMAIN 164 185 POLY-SER.
CC  FT  DOMAIN 205 218 POLY-THR.

```



QY 184 STSSSSSSSS 193  
 Db 9 STSSSSSSSS 18

RESULT 7  
 UBC6\_MOUSE STANDARD: PRT: 193 AA.  
 ID UBC6\_MOUSE  
 AC P52482;  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE UBIQUITIN-CONJUGATING ENZYME E2-21 KDA UBC6 (EC 6.3.2.19) (UBIQUITIN-  
 DE PROTEIN LIGASE) (UBIQUITIN CARRIER PROTEIN).  
 GN UBC6 OR UBCM3 OR UBC5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-Brain;  
 RX MEDLINE=96162026; PubMed=8576256.  
 RA Matuschewski K., Hauser H.P., Tietler M., Jentsch S.;  
 RT "Identification of a novel family of ubiquitin-conjugating enzymes  
 RT with distinct amino-terminal extensions.";  
 RL J. Biol. Chem. 271:2789-2794(1996).  
 CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER  
 CC PROTEINS. MEDIANES THE SELECTIVE DEGRADATION OF SHORT-LIVED AND  
 CC ABNORMAL PROTEINS.  
 CC -1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE =  
 CC AMP + PYROPHOSPHATE + PROTEIN N-UBIQUITYLYSINE.  
 CC -1- PATHWAY: SECOND STEP IN UBIQUITIN CONJUGATION.  
 CC -1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR  
 CC UBIQUITIN-THIOLESTER FORMATION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.  
 CC STRONGEST, TO DROSOPHILA UBCD2.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X92665; CAA63353.1; -  
 DR HSP; P15731.2UC.  
 DR MGD; MG1:107411; Ubc65.  
 DR InterPro; IPR000608; UQ\_conjugat.  
 DR Pfam; PF00179; UQ\_con; 1.  
 DR SMART; SM00212; UBC; 1.  
 DR PROSITE; PS00183; UBIQUITIN\_CONJUGAT\_1; 1.  
 DR PROSITE; PS50127; UBIQUITIN\_CONJUGAT\_2; 1.  
 DR Ubiqultin conjugation; Ligase; Multigene family.  
 FT DOMAIN 9 18 POLY-SER.  
 FT BINDING 131 131 UBIQUITIN (BY SIMILARITY).  
 SQ SEQUENCE 193 AA; 21333 MW; B535B3095EFC445 CRC64;

QY 184 STSSSSSSSS 193  
 Db 9 STSSSSSSSS 18

RESULT 8  
 AXIB\_ARATH STANDARD: PRT: 246 AA.  
 ID AXIB\_ARATH  
 AC Q38829;

Query Match 1.68; Score 10; DB 1; Length 193;  
 Best Local Similarity 100.0%; Pred. No. 0.36;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE AUXIN-RESPONSIVE PROTEIN IAA1 (INDOLEACETIC ACID-INDUCED PROTEIN 11).  
 GN IAA1 OR AT4G28640 OR T5F17.90.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RX MEDLINE=95387393; PubMed=7658471;  
 RA Abel S., Nguyen M.D., Theologis A.;  
 RT "The PS-IAA4/5-like family of early auxin-inducible mRNAs in  
 RT Arabidopsis thaliana.";  
 RL J. Mol. Biol. 251:533-549(1995).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, COLUMBIA;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Voickaert G.,  
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,  
 RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,  
 RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,  
 RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,  
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.-A., McCullagh B., Blahm L., Robben J.,  
 RA Van der Schueren J., Gymonpretz B., Chuang Y.-J., Vandenbussche F.,  
 RA Breken M., Weltjens I., Voet M., Bastlaens I., Aert R., Defoor E.,  
 RA Weltzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dikse W.,  
 RA Moeljan P., Klein lanthorst R., Rose M., Hauf J., Koetter P.,  
 RA Berner S., Hempel S., Feldpausch M., Lambertz S., Van den Biele H.,  
 RA De Keyser A., Buysschaert C., Gielen J., Villarroel R., De Clercq R.,  
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,  
 RA Petrett A., Ralndream M.-A., Lyne M., Benes V., Rechmann S.,  
 RA Boktor D., Bloeker H., Scharte M., Grimm M., Loehert T.-H.,  
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Grandpré K., Dauner D., Herzl A.,  
 RA Neumann S., Argitlon A., Vitale D., Liqouri R., Piravandi E.,  
 RA Massenet O., Quigley F., Clapaud G., Muendlein A., Felber R.,  
 RA Schnabel S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
 RA Chefford F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,  
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,  
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Hebermann K.,  
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
 RA Latreille T., Kallick J., Graves T., Harmon G., Edwards J.,  
 RA Llorenç L., Courtney L., Clout J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
 RA Nelson J., Spieth J., Ryan E., Andrews S., Gelsel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
 RA Ma P., Zhong J., Preston R., Vil D., Shekhar M., Matero A., Shah R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,  
 RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lochl M., Johnson A.,  
 RA Chen E., Marra M., Martienssen R., McComble W.R.;  
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 402:769-777(1999).  
 CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR  
 CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL  
 CC GROWTH (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- INDUCTION: BY AUXIN.



```

CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL: U18413; AAC49052.1; -
CC EMBL: AL161573; CAB81452.1; -
CC EMBL: 7127; Atatrh.1524;7127.
CC InterPro: IPR003311; AUX_IAA.
CC Pfam: PF02309; AUX_IAA; 1.
CC Multigene family: Nuclear protein; Translation regulation.
CC DOMAIN 57 67
CC SEQUENCE 246 AA; 26516 MW; 449054E312FBBDAD CRC64;
CC -----
CC Query Match 1.6%; Score 10; DB 1; Length 246;
CC Best Local Similarity 100.0%; Pred. No. 0.44;
CC Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC OY 186 SSSSSSSSS 195
CC Db 57 SSSSSSSSS 66
CC -----
CC RESULT 9
CC LYTE_BACSU STANDARD; PRT; 334 AA.
CC ID LYTE_BACSU
CC AC P54421;
CC DT 01-OCT-1996 (Rel. 34, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DT 20-AUG-2001 (Rel. 40, Last annotation update)
CC DE PUTATIVE ENDOPEPTIDASE LYTE PRECURSOR (PHOSPHATASE-ASSOCIATED PROTEIN
CC DE PAPO) (CELL WALL-ASSOCIATED POLYPEPTIDE CMBP33).
CC OS LYTE OR PAPO.
CC GN Bacillus subtilis.
CC OC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC OC Bacillus/Staphylococcus group; Bacillus.
CC NCBI_TaxID=1423;
CC [1]
CC RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-36.
CC RC STRAIN=168;
CC MEDLINE=98117063; PubMed=9457885;
CC RX Margot P., Mahlen M., Gholamhusein A., Pigot P., Karamata D.;
CC RA "The lyte gene of Bacillus subtilis 168 encodes a cell wall
CC RT hydrolase.";
CC RL J. Bacteriol. 180:749-752(1998).
CC [2]
CC RP SEQUENCE OF 311-334 FROM N.A.
CC RC STRAIN=168 / SMY;
CC RX MEDLINE=94321340; PubMed=8045898;
CC RA Jin S., Sonenshein A.L.;
CC RT "Identification of two distinct Bacillus subtilis citrate synthase
CC RL genes.";
CC RL J. Bacteriol. 176:4669-4679(1994).
CC -1- FUNCTION: PUTATIVE D-GLUTAMATE - M-DIAMINOPIMLATE ENDOPEPTIDASE.
CC CELL WALL HYDROLASE INVOLVED IN CELL AUTOLYSIS.
CC -1- SUBCELLULAR LOCATION: CELL-WALL BOUND.
CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
CC BINDING.
CC -1- SIMILARITY: BELONGS TO THE E-COLI NLPC / LISTERIA P60 FAMILY.
CC -1- SIMILARITY: CONTAINS 3 LYSM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)

```

```

CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL: U38819; AAC25975.1; -
CC EMBL: U05256; -, NOT_ANNOTATED_CDS.
CC Subtilist; BG11406; LYTE.
CC InterPro: IPR002482; LYSM.
CC InterPro: IPR000064; NLPC_P60.
CC Pfam: PF01476; LYSM; 3.
CC Pfam: PF00877; NLPC_P60; 1.
CC SMART: SM00257; LYSM; 3.
CC Cell wall; Hydrolase; Repeat; Signal; Complete proteome.
CC KW SMART; SM00257; LYSM; 3.
CC FT SIGNAL 1 25
CC FT CHAIN 26 334
CC FT REPEAT 28 70
CC FT REPEAT 88 130
CC FT REPEAT 151 193
CC FT DOMAIN 73 86
CC FT DOMAIN 132 141
CC FT DOMAIN 144 149
CC FT DOMAIN 204 211
CC SEQUENCE 334 AA; 35455 MW; D809C7DD4BAC475D CRC64;
CC -----
CC Query Match 1.6%; Score 10; DB 1; Length 334;
CC Best Local Similarity 100.0%; Pred. No. 0.57;
CC Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC OY 184 STSSSSSSSS 193
CC Db 132 STSSSSSSSS 141
CC -----
CC RESULT 10
CC PLTB_PHAVU STANDARD; PRT; 461 AA.
CC ID PLTB_PHAVU
CC AC 043822;
CC DT 15-DEC-1998 (Rel. 37, Created)
CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
CC DT 15-DEC-1998 (Rel. 37, Last annotation update)
CC DE GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE PRECURSOR (EC 2.3.1.15) (GPAT).
CC OS Phaseolus vulgaris (Kidney bean) (French bean).
CC GN PLTB.
CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
CC NCBI_TaxID=3885;
CC [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=CV. ANNABEL; TISSUE=leaf;
CC RX MEDLINE=95232196; PubMed=7716242;
CC RA Fritz M., Heinz E., Wolter F.P.;
CC RT "Cloning and sequencing of a full-length cDNA coding for
CC RL sn-glycerol-3-phosphate acyltransferase from Phaseolus vulgaris.";
CC RL Plant Physiol. 107:1039-1040(1995).
CC -1- FUNCTION: ESTERIFIES ACYL-GROUP FROM ACYL-ACP TO THE SN-1 POSITION
CC OF GLYCEROL-3-PHOSPHATE. THE ENZYME FROM CHILLING-RESISTANT PLANTS
CC DISCRIMINATE AGAINST NON-FLUID PALMITIC ACID AND SELECT OLEIC ACID
CC WHEREAS THE ENZYME FROM SENSITIVE PLANTS ACCEPTS BOTH FATTY ACIDS.
CC -1- CATALYTIC ACTIVITY: ACYL-COA + SN-GLYCEROL 3-PHOSPHATE =
CC COA + 1-ACYL-SN-GLYCEROL 3-PHOSPHATE.
CC -1- PATHWAY: FIRST STEP IN DE NOVO PHOSPHOLIPID BIOSYNTHESIS. IT MAY
CC ALSO FUNCTION IN THE REGULATION OF MEMBRANE BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST STROMA.
CC -1- SIMILARITY: BELONGS TO THE PLANT GPAT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)

```

DR EMBL: X79722; CAA56159.1; -  
 DR InterPro: IPR002123; Acyltransferase.  
 KW Phospholipid biosynthesis; Transferase; Acyltransferase;  
 KW TRANSIT 1 96 CHLOROPLAST (POTENTIAL).  
 FT CHAIN 97 461 GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE.  
 SQ SEQUENCE 461 AA; 50697 MW; EAC4FC837908B38A CRC64;

Query Match 1.6%; Score 10; DB 1; Length 461;  
 Best Local Similarity 100.0%; Pred. No. 0.74;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 TSSSSSSSS 192  
 DB 49 TSSSSSSSS 58

RESULT 11  
 ID KG3H\_DICDI STANDARD: PRT: 468 AA.  
 AC P51136:  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE GLYCOCEN SYNTHASE KINASE-3 HOMOLOG (EC 2.7.1.1) (GSK-3).  
 GN GSKA.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX2;  
 RX MEDLINE=95112340; PubMed=7813009;  
 RA Hatwood A.J., Plyte S.E., Woodgett J., Strutt H., Kay R.R.;  
 RL "Glycogen synthase kinase 3 regulates cell fate in Dictyostelium";  
 Cell 80:139-148(1995).  
 CC -1- FUNCTION: DURING CELLULAR DIFFERENTIATION IT MAY MEDIATE A CAMP  
 STIMULATED SIGNAL TRANSDUCTION PATHWAY THAT REGULATES PRESORE AND  
 PRESTALK B CELL PROPORTIONS.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC CDC2/CDKX SUBFAMILY; GSK-3 SUBSUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L34674; AAA65968.1; -  
 DR HSSP: P24941; IAO1.  
 DR Dictydb; DD01073; gskA.  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_kin\_actsite.  
 DR Pfam; PF00069; pkinase; 1.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW phosphorylation.  
 FT DOMAIN 57 340 PROTEIN KINASE.  
 FT NP\_BIND 63 71 ATP (BY SIMILARITY).  
 FT BINDING 86 86 ATP (BY SIMILARITY).  
 FT ACT\_SITE 180 180 BY SIMILARITY.  
 FT MOD\_RES 215 215 PHOSPHORYLATION (BY SIMILARITY).  
 FT DOMAIN 23 28 POLY-THR.  
 FT DOMAIN 29 38 POLY-SER.  
 FT DOMAIN 405 410 POLY-SER.  
 FT DOMAIN 429 433 POLY-SER.  
 FT DOMAIN 456 465 POLY-THR.

SQ SEQUENCE 468 AA; 51598 MW; 50DCE4B7D601D5F9 CRC64;

Query Match 1.6%; Score 10; DB 1; Length 468;  
 Best Local Similarity 100.0%; Pred. No. 0.75;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 TSSSSSSSS 194  
 DB 28 TSSSSSSSS 37

RESULT 12  
 ID YOD0\_YEAST STANDARD: PRT: 484 AA.  
 AC Q08193:  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHETICAL 51.9 KDA PROTEIN IN MSEL-LAG2 INTERGENIC REGION  
 DE PRECURSOR.  
 GN YOL030W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Habbig B., Hattenhorst U., Hollenberg C.P., Ramezani Rad M.;  
 RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE GAS1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z74772; CAA99030.1; -  
 DR SGD: S0005390; YOL030W.  
 DR KW Hypothetical protein; Glycoprotein; Signal.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 484 HYPOTHETICAL PROTEIN YOL030W.  
 FT DOMAIN 395 403 ASP/GLU-RICH (HIGHLY ACIDIC).  
 FT DOMAIN 405 460 SER-RICH.  
 FT DOMAIN 405 427 POLY-SER.  
 FT DOMAIN 432 439 POLY-SER.  
 FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 484 AA; 51869 MW; E65B9F4689BBB34C CRC64;

Query Match 1.6%; Score 10; DB 1; Length 484;  
 Best Local Similarity 100.0%; Pred. No. 0.77;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 TSSSSSSSS 194  
 DB 404 TSSSSSSSS 413

RESULT 13  
 ID Y859\_TREPA STANDARD: PRT: 494 AA.  
 AC O83831; O83832;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL PROTEIN TP0859/TP0860 PRECURSOR.  
GN TP0859/TP0860.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=160;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NICHOIS;  
RX MEDLINE=98332770; PubMed=9665876;  
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utechtack T.,  
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
RA Venter J.C.;  
RT "Complete genome sequence of Treponema pallidum, the syphilis  
RT spirochete."  
RL Science 281:375-388(1998).  
CC -1- SIMILARITY: BELONGS TO THE UPF0164 FAMILY.  
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. REF.1 SEQUENCE DIFFERS  
CC FROM THAT SHOWN DUE TO A FRAMESHIFT IN POSITION 232.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AE001256; AAC65831.1; ALT\_FRAME.  
DR EMBL; AE001256; AAC65832.1; ALT\_FRAME.  
DR TIGR; TP0859; .  
DR TIGR; TP0860; .  
KM Hypothetical protein; Signal; Complete proteome.  
FT SIGNAL 1 44 POTENTIAL.  
FT CHAIN 45 494 HYPOTHETICAL PROTEIN TP0859/TP0860.  
FT SEQUENCE 494 AA; 52883 MW; 5B419C60CEAF840 CRC64;  
-----  
Query Match 1.6%; Score 10; DB 1; Length 494;  
Best Local Similarity 100.0%; Pred. No. 0.79;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
OY 8 AGSGGGGGG 17  
DB 296 AGSGGGGGG 305  
-----  
RESULT 14  
YN23\_YEAST STANDARD; PRT; 503 AA.  
AC P53832;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOTHETICAL 52.3 KDA PROTEIN IN MRPL10-ERG24 INTERGENIC REGION  
DE PRECURSOR.  
GN YNL283C OR N0583.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
[1]  
RN SEQUENCE FROM N.A.  
RP Mesenquy F., Dubois E., Vierendeels F., Scherens B., Pierard A.,  
RA Glansdorff N.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO S.POMBE SPB305.14C.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; 271559; CAA96195.1; .  
DR SGD; S0005227; YNL283C.  
DR InterPro; IPR002889; MSC.  
DR Pfam; PF01822; MSC; 1.  
DR SMART; SM00321; MSC; 1.  
KM Hypothetical protein; Transmembrane; Signal.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 503 HYPOTHETICAL PROTEIN YNL283C.  
FT TRANSMM 61 81 POTENTIAL.  
FT TRANSMEM 326 346 POTENTIAL.  
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 429 429 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 475 475 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT SEQUENCE 503 AA; 52292 MW; F2392A73C5CBAB50 CRC64;  
-----  
Query Match 1.6%; Score 10; DB 1; Length 503;  
Best Local Similarity 100.0%; Pred. No. 0.8;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
OY 185 TSSSSSSSS 194  
DB 182 TSSSSSSSS 191  
-----  
RESULT 15  
MS2P\_HUMAN STANDARD; PRT; 519 AA.  
ID MS2P\_HUMAN  
AC O43462;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE STEROL-REGULATORY ELEMENT-BINDING PROTEIN INTRAMEMBRANE PROTEASE  
DE (EC 3.4.24.-) (SITE-2 PROTEASE).  
GN S2P.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Endothelial;  
RX MEDLINE=98324087; PubMed=9659902;  
RA Rawson R.B., Zelenski N.G., Nijhawan D., Ye J., Sakai J., Hasan M.T.,  
RA Chang T.Y., Brown M.S., Goldstein J.L.;  
RT "Complementation cloning of S2P, a gene encoding a putative  
RT metalloprotease required for intramembrane cleavage of SREBPs."  
RL Mol. Cell 1:47-57(1997).  
CC -1- FUNCTION: INTRAMEMBRANE PROTEOLYSIS OF STEROL-REGULATORY ELEMENT-  
CC BINDING PROTEINS (SREBPs) WITHIN THE FIRST TRANSMEMBRANE SEGMENT  
CC THEREBY RELEASING THE N-TERMINAL SEGMENT WITH A PORTION OF THE  
CC TRANSMEMBRANE SEGMENT ATTACHED. SITE-2 CLEAVAGE COMES AFTER SITE-1  
CC CLEAVAGE WHICH TAKE PLACES IN THE LOWER LOOP.  
CC -1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG,  
CC LIVER, MUSCLE, KIDNEY AND PANCREAS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50 (ZINC  
CC METALLOPROTEASE).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AF019612; AAC51937.1; -.  
 DR MEROPS; M50.001; -.  
 DR InterPro; IPR001193; Peptidase\_M50.  
 DR InterPro; IPR000130; Zn\_MTpeptidse.  
 DR Pfam; PF02163; Peptidase\_M50; 1.  
 DR PRINTS; PR01000; SREBPS2PTASE.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolyase; Metalloprotease; Zinc; Transmembrane.  
 FT METAL 171 171 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 172 172 BY SIMILARITY.  
 FT METAL 175 175 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT TRANSMEM 4 24 POTENTIAL.  
 FT TRANSMEM 75 95 POTENTIAL.  
 FT TRANSMEM 186 206 POTENTIAL.  
 FT TRANSMEM 229 249 POTENTIAL.  
 FT TRANSMEM 448 468 POTENTIAL.  
 FT TRANSMEM 493 513 POTENTIAL.  
 FT DOMAIN 109 136 POLY-SER.  
 FT DOMAIN 380 384 POLY-SER.  
 FT MUTAGEN 171 171 H->F: ACTIVITY ABOLISHED.  
 FT MUTAGEN 172 172 E->A: ACTIVITY ABOLISHED.  
 FT MUTAGEN 172 172 E->Q: ACTIVITY ABOLISHED.  
 FT MUTAGEN 172 172 E->D: ACTIVITY PARTIALLY ABOLISHED.  
 FT MUTAGEN 175 175 H->F: ACTIVITY ABOLISHED.  
 SQ SEQUENCE 519 AA; 57443 MW; 247D69EDFD7747BD CRC64;

Query Match 1.6%; Score 10; DB 1; Length 519;  
 Best Local Similarity 100.0%; Pred. No. 0.82;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 186 SSSSSSSSSL 195  
 |||||||||  
 Db 128 SSSSSSSSSL 137

Search completed: January 22, 2002, 16:37:08  
 Job time: 318 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:31:00 ; Search time 40.87 Seconds  
(without alignments)  
2229.096 Million cell updates/sec

Title: US-09-485-529-7

Perfect score: 623  
Sequence: 1 MKREYDAGSGGGGGGMSGMS.....TIGMHTPLTATSAWRLAGP 623

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL\_17:\*

- 1: sp.archaea:\*
- 2: sp.bacteria:\*
- 3: sp.fungi:\*
- 4: sp.human:\*
- 5: sp.invertebrate:\*
- 6: sp.mammal:\*
- 7: sp.mbc:\*
- 8: sp.organelle:\*
- 9: sp.phage:\*
- 10: sp.plant:\*
- 11: sp.protoent:\*
- 12: sp.virus:\*
- 13: sp.vertebrate:\*
- 14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	623	100.0	623	10 Q9ST59	Q9ST59 titilium ae
2	132	21.2	620	10 Q9ST48	Q9ST48 zea mays (m
3	89	14.3	625	10 Q9MB96	Q9MB96 oryza sativ
4	23	3.7	547	10 Q9SRP9	Q9SRP9 arabidopsis
5	22	3.5	511	10 Q9C8Y3	Q9C8Y3 arabidopsis
6	22	3.5	662	10 Q65367	Q65367 arabidopsis
7	19	3.0	493	10 Q9AS97	Q9AS97 oryza sativ
8	19	3.0	523	10 Q9LF53	Q9LF53 arabidopsis
9	19	3.0	532	10 Q23643	Q23643 arabidopsis
10	19	3.0	532	10 Q23724	Q23724 arabidopsis
11	19	3.0	533	10 Q9LQ78	Q9LQ78 arabidopsis
12	19	3.0	587	10 Q23725	Q23725 arabidopsis
13	19	3.0	587	10 Q23642	Q23642 arabidopsis
14	19	3.0	587	10 Q9SLH3	Q9SLH3 arabidopsis
15	16	2.6	275	10 Q43476	Q43476 hordeum vul
16	12	1.9	118	5 Q9U1K5	Q9U1K5 drosophila
17	12	1.9	119	5 Q9V401	Q9V401 drosophila
18	12	1.9	153	3 Q08906	Q08906 saccharomyc
19	12	1.9	229	11 Q9JM93	Q9JM93 mus musculu

20	12	1.9	390	2 Q9PHR0	Q9PHR0 campylobact
21	12	1.9	409	3 Q9HED2	Q9HED2 neurospora
22	12	1.9	446	3 Q9Y849	Q9Y849 kluyveromyc
23	12	1.9	574	4 Q9BT08	Q9BT08 homo sapien
24	12	1.9	856	5 Q9GZ10	Q9GZ10 drosophila
25	12	1.9	989	5 Q9W2S4	Q9W2S4 drosophila
26	12	1.9	1297	4 Q15025	Q15025 homo sapien
27	12	1.9	1383	12 Q89232	Q89232 canine herp
28	12	1.9	2161	12 Q9E1P1	Q9E1P1 meleagrid h
29	12	1.9	2164	12 Q9DH52	Q9DH52 meleagrid h
30	12	1.8	164	5 Q19817	Q19817 caenorhabdi
31	11	1.8	187	11 Q99NV0	Q99NV0 hydrochoeru
32	11	1.8	193	11 Q99NV1	Q99NV1 cavia teschu
33	11	1.8	216	10 Q9C9J8	Q9C9J8 arabidopsis
34	11	1.8	238	5 Q44958	Q44958 caenorhabdi
35	11	1.8	291	10 Q9M160	Q9M160 arabidopsis
36	11	1.8	318	10 Q9S782	Q9S782 oryza sativ
37	11	1.8	321	5 Q9W3Z5	Q9W3Z5 drosophila
38	11	1.8	371	10 Q9FN34	Q9FN34 arabidopsis
39	11	1.8	417	5 Q9XY07	Q9XY07 lytechinus
40	11	1.8	429	10 Q9C515	Q9C515 arabidopsis
41	11	1.8	450	10 Q9DA19	Q9DA19 mus musculu
42	11	1.8	575	10 Q23083	Q23083 arabidopsis
43	11	1.8	600	13 Q90237	Q90237 anolis pulc
44	11	1.8	614	5 P91720	P91720 drosophila
45	11	1.8	632	10 Q80569	Q80569 arabidopsis

#### ALIGNMENTS

RESULT	ID	Q9ST59	PRELIMINARY:	PRT:	623 AA.
Q9ST59	AC	Q9ST59;			
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)			
DE	GIBBERELLIN RESPONSE MODULATOR.				
GN	RHT-DIA.				
OS	Triticum aestivum (wheat).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;				
OC	Triticeae; Triticum.				
OX	NCBI_TaxID=4565;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=99347734; PubMed=10421366;				
RA	Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,				
RA	Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F.,				
RA	Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;				
RT	"Green revolution" genes encode mutant gibberellin response				
RT	modulators."				
RL	Nature 400:256-261(1999).				
DR	EMBL: AJ242531; CAB51555.1; .				
SQ	SEQUENCE 623 AA; 65337 MW; F2BAC3496D0A84F CRC64;				
Query Match 100.0%; score 623; DB 10; length 623;					
Best local similarity 100.0%; Pred. No. 0;					
Matches 623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MKREYDAGSGGGGGGMSGSSDDKMMVSAAGGEEVDELLAAGYKVRASMDADVAOKI 60			
DB	1	MKREYDAGSGGGGGGMSGSSDDKMMVSAAGGEEVDELLAAGYKVRASMDADVAOKI 60			
QY	61	EOLEAMMGGVGAAPDSEFATLATDTVHYNPDDLSWVESMSELNAPPPPLPPAP 120			
DB	61	EOLEAMMGGVGAAPDSEFATLATDTVHYNPDDLSWVESMSELNAPPPPLPPAP 120			
QY	121	QLNAASSVTYSGGYFDLPSPVSSSSSYIALRIPSPAGATAPADLSADVDPKRMRT 180			
DB	121	QLNAASSVTYSGGYFDLPSPVSSSSSYIALRIPSPAGATAPADLSADVDPKRMRT 180			

OY	181	GGTSSSSSSSSSLGGGAASSVVEAAPPAAANATPALPVVVDTQEGITLVHALLA	240
Db	181	GGSTSSSSSSSSSLGGGARSVVEAAPVAAAAANATPALPVVVDTQEGILVHALLA	240
OY	241	CAEAVOENLSAEALVKQIPLPLAASOGAMRKVAAFEGELARRYFERPQDSSILDA	300
Db	241	CAEAVOENLSAAALVKQIPLPLAASOGAMRKVAAYFEGELARRVFRRPQDSSILDA	300
OY	301	AFADLLAHAFESCPYLKFAHTANAQLLEAFAGCRVHVVDGIGKOMWPLLQALAI	360
Db	301	AFAOLLHAHFESCPYLKFHFTANOAILLEAFAGCRRVHVVDGIGKOMWPALLQALAI	360
OY	361	RPGPPSRRLTGVPPOPDETDALQQYGMLAQFAHTIRVDFQGLVATLADLEPFMI	420
Db	361	RPGPPSRRLTGVPPOPDETDALQQYGKMLAQFAHTIRVDFQGLVATLADLEPFMI	420
OY	421	QPGEEDPNNEPEVIYAIVSVFEMHRLLAQBGALEKVLGTVAVRPRIVTVVEQEANHNSG	480
Db	421	QPGEEDPNNEPEVIYAIVSVFEMHRLLAQBGALEKVLGTVAVRPRIVTVVEQEANHNSG	480
OY	481	TFLDRFTESLHYISTMDDSLGSGSGGGSPDEFVSSSGAAAAPAAAGTOVMSEYVLGHOICN	540
Db	481	TFLDRFTESLHYISTMDPSLEGSSGGGSPDEFVSSSGAAAAPAAAGTOVMSEYVLGHQICN	540
OY	541	VVACEGAEERTERHETTLQOMNRRLCNAGFEFVHLGSNNYKAOSTLLALFAGGDGYKYEEKE	600
Db	541	VVACEGAEERTERHETTLQOMNRRLCNAGFEFVHLGSNNYKASTLLALFAGGDGYKYEEKE	600
OY	601	GCLTLGWHTRPLITSAMRLAPG	623
Db	601	GCLTLGWHTRPLITSAMRLAPG	623

RESULT	2			
Q9ST48				
ID	Q9ST48	PRELIMINARY;	PRT;	630 AA.
AC	Q9ST48;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	GIBBERELLIN RESPONSE MODULATOR (FRAGMENT).			
DE	D8.			
OS	Zea mays (Maize).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade			
OC	Panicoidae; Andropogonae; Zea.			
XC	NCBI_TaxID=4577;			
XC	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=99347734; PubMed=10421366;			
RA	Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,			
RA	Flintham J.E., Beales J., Fish L.U.J., Worland A.J., Pelica F.,			
RA	Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;			
RT	"Green revolution" genes encode mutant gibberellin response			
RT	modulators.";			
RL	Nature 400:256-261(1999).			
DR	EMBL; AJ242530; CAB51557.1; .			
FT	NON_TER	630	630	
Q0	SEQUENCE	630 AA;	66028 MW;	3D56851726C51042 CRC64;

	Query Match	Best Local Similarity	21.2%; Score 132	DB 10-115; length 630;
			100.0%; Pred. No. 1	Be-115; length 630;
	Matches 132;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0;
QY	293	PDSLLDAAAFADLHAHYESCPYIKFAHFTANQAILLEAFACGRVHYVDFSIKGMQMP	352	
Db	302	PDSLLDAAAFADLHAHYESCPYIKFAHFTANQAILLEAFACGRVHYVDFSIKGMQMP	361	
QY	353	ALLQALALRPGSPFSFRLTGVGPDPDETDAQQVGMWLAEQFAHTRIVDFQGRGLVAATL	412	
Db	362	ALLQALALRPGSPFSFRLTGVGPDPDETDAQQVGMWLAEQFAHTRIVDFQGRGLVAATL	421	

```
QY      413 ADLEPFMLQPEG 424
          |||||
Db      422 ADLEPFMLQPEG 433
```

	RESULT	3
Q9MB96		
ID	Q9MB96	PRELIMINARY; PRT; 625 AA.
AC	Q9MB96	
DT	01-OCT-2000 (TREMBLrel. 15, Created)	
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)	
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)	
DE	OSGAI.	
GN	OSGAI.	
OS	Oryza sativa (Rice).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
OC	Erihaltoideae; Oryzeae; Oryza.	
OX	NCBI_TaxID=4530;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CV. NIPPONBARE;	
RX	MEDLINE=20179680; PubMed=10713441;	
RA	Ogawa M., Kusano T., Katsumi M., Sano H.;	
RT	"Rice gibberellin-insensitive gene homolog, OSGAI, encodes a nuclear-	
RT	localized protein capable of gene activation at transcriptional	
RT	level.";	
RL	Gene 245:21-29(2000).	
DR	EMBL; AB030956; BAA90749.1; -	
SO	SEQUENCE 625 AA; 65406 MW; 034FF02719D42E97 CRC64;	

	Query Match	14.3%	Score 89;	DB 10;	Length 625;
	Best Local Similarity	100.0%	Pred. No. 5.6e-75;		
Matches	89;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	337	RHHVHDEGKQOMRPA	LQALATATARRGGRSPRLTGVGRPDDEHDA	LQOYVWKLQAFH	396
DB	343	RHHVHDEGKQOMRPA	LQALATATARRGGRSPRLTGVGRPDDEHDA	LQOYVWKLQAFH	402
QY	397	TIRVDFQYRGILVAAT	LADLERPMLOPEGE	425	
DB	403	TIRVDFQYRGILVAAT	LADLERPMLOPEGE	431	

RESULT	4		
Q9SRP9			
Q9SRP9	PRELIMINARY;	PRT;	547 AA.
Q9SRP9;			
DT 01-MAY-2000	(TREMBLrel. 13, Created)		
DT 01-MAY-2000	(TREMBLrel. 13, Last sequence update)		
DT 01-MAY-2000	(TREMBLrel. 13, Last annotation update)		
DE	RGAI-LIKE PROTEIN.		
GN	T21P5.13.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Euraryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eucotsids II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RI	[1]		
RM	SEQUENCE FROM N.A.		
RP	STRAIN=CV. COLUMBIA;		
RC	Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,		
RA	Ronning C.M., Koo H., Fujii C.Y., Uteback T.R., Barnstead M.E.,		
RA	Bowman C.L., White O., Nierman W.C., Fraser C.M.;		
RT	"Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence.";		
RL	Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.		
RS	EMBL; AC009895; AAF01590.1; -		
SO	SEQUENCE	547 AA;	60493 MW; C4D18D5951D95634 CRC64;

Query Match	3.7%;	Score 23;	DB 10;	Length 547;
-------------	-------	-----------	--------	-------------



```

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
RA Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL391150; CAC01893.1; -.
SQ SEQUENCE 523 AA; 57326 MW; 0FCE0BD13403C35 CRC64;

Query Match 3.0%; Score 19; DB 10; Length 532;
Best Local Similarity 100.0%; Pred. No. 4.1e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 313 SCPYKFAHFTANQALIEA 331
DB 234 SCPYKFAHFTANQALIEA 252

RESULT 9
OZ3643 PRELIMINARY; PRT; 532 AA.
AC OZ3643;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE RGA2 PROTEIN.
GN RGA2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97379310; PubMed=9237632;
RA Truong H.N., Caboche M., Daniel-Vedele F.;
RA "Sequence and characterization of two Arabidopsis thaliana CDNAs
RT isolated by functional complementation of a yeast gln3 gdh1 mutant.";
RA FEBS Lett. 410:213-218(1997).
DR EMBL; Y11337; CAA72178.1; -.
DR Mendel; 24146; Arabidopsis; 17AC719CA8072239 CRC64;
SQ SEQUENCE 532 AA; 58789 MW; 17AC719CA8072239 CRC64;

Query Match 3.0%; Score 19; DB 10; Length 532;
Best Local Similarity 100.0%; Pred. No. 4.2e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALIEAF 332
DB 246 CPYLKFAHFTANQALIEAF 264

RESULT 10
OZ3724 PRELIMINARY; PRT; 532 AA.
AC OZ3724;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE GAI PROTEIN.
GN GAI.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LANDSBERG_ERECTA;
RX MEDLINE=98051192; PubMed=9389651;
RA Peng J., Carol P., Richards D.E., King K.E., Cowling R.J.,
RA Murphy G.P., Harberd N.P.;
RT "The Arabidopsis GAI gene defines a signaling pathway that negatively
RT regulates gibberellin responses.";
RL Genes Dev. 11:3194-3205(1997).
DR EMBL; Y15193; CAA5492.1; -.
DR Mendel; 24070; Arabidopsis; 4AF4BC6EC4265503 CRC64;
SQ SEQUENCE 532 AA; 58789 MW; 4AF4BC6EC4265503 CRC64;

Query Match 3.0%; Score 19; DB 10; Length 532;
Best Local Similarity 100.0%; Pred. No. 4.2e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALIEAF 332
DB 246 CPYLKFAHFTANQALIEAF 264

RESULT 11
OZ1078 PRELIMINARY; PRT; 533 AA.
AC OZ1078;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE F10B6.34.
GN Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Chao O., Shinn P., Dunn P., Buehler E., Kahn S., Kim C., Walker M.,
RA Williams S., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
RA Hansen N.F., Hulzar L., Kremetska I., Lenz C., Li J., Liu S.,
RA Lueros S., Rowley D., Schwartz J., Toriumi M., Vysotskaya V., Yu G.,
RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome
RT 1.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao O., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Muharsy N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Theologis A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006917; AAF79228.1; -.
SQ SEQUENCE 533 AA; 58926 MW; 6464B8C129D79528 CRC64;

Query Match 3.0%; Score 19; DB 10; Length 532;
Best Local Similarity 100.0%; Pred. No. 4.2e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALIEAF 332

```



Db 247 CPYLKFAHFTANQALILEAF 265

RESULT 12  
ID 023725 PRELIMINARY; PRT: 587 AA.

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE GRS PROTEIN.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Peng J., Carol P., Richards D.E., King K.E., Cowling R.J.,  
RA Murphy G.P., Harberd N.P.;  
RL Genes Dev. 0:0-0(0).  
DR EMBL: Y15194; CAA75493.1; -  
DR Mendel: 24071; Atarch:3051;24071.  
SQ SEQUENCE 587 AA; 64006 MW; F6F6C7738E7DCA9 CRC64;

Query Match  
Best Local Similarity 3.0%; Score 19; DB 10; Length 587;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALILEAF 332  
Db 299 CPYLKFAHFTANQALILEAF 317

RESULT 13  
ID 023642 PRELIMINARY; PRT: 587 AA.

AC 023642;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE RGL1 PROTEIN.  
GN RGL1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=97379310; PubMed=9237632;  
RA Truong H.N., Caboche M., Daniel-Vedele F.;  
RT "Sequence and characterization of two Arabidopsis thaliana cDNAs  
RT isolated by functional complementation of a yeast gln3 gdh1 mutant.";  
RL FEBS Lett. 410:213-218(1997).  
DR EMBL: Y11336; CAA72177.1; -  
DR Mendel: 24145; Atarch:3051;24145.  
SQ SEQUENCE 587 AA; 64023 MW; DDA7A3C741FB51EF CRC64;

Query Match  
Best Local Similarity 3.0%; Score 19; DB 10; Length 587;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALILEAF 332  
Db 299 CPYLKFAHFTANQALILEAF 317

RESULT 14

O9SLH3  
ID O9SLH3 PRELIMINARY; PRT: 587 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE PUTATIVE RGL1, GIBBERELLIN RESPONSE MODULATION PROTEIN.  
GN AT2G01570.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., Vanden S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";  
RL Nature 402:761-768(1999).  
DR EMBL: AC005560; AAC67333.1; -  
DR Mendel: 587 AA; 64035 MW; FC92E7F9408072AA CRC64;  
SQ SEQUENCE 587 AA; 64035 MW; FC92E7F9408072AA CRC64;

Query Match  
Best Local Similarity 3.0%; Score 19; DB 10; Length 587;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALILEAF 332  
Db 299 CPYLKFAHFTANQALILEAF 317

RESULT 15  
ID 043476 PRELIMINARY; PRT: 275 AA.

AC 043476;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE MAMMALIAN ACYL COA OXIDASE HOMOLOGOUS (FRAGMENT).  
GN CDR29.  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Triticeae; Hordeum.  
OX NCBI\_TaxID=4513;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. ARDA; TISSUE=GREEN LEAF;  
RA Grossi M., Galli M., Stanca A.M., Cattivelli L.;  
RL Plant Sci. 105:71-80(1995).  
DR EMBL: X84055; CAA58874.1; -  
DR Mendel: 12863; Horvu:2063;12863.  
DR InterPro: IPR002655; ACOX.  
DR Pfam: PF01756; ACOX; 1.  
FT NON\_TER 1  
SQ SEQUENCE 275 AA; 30689 MW; 332811FD05827472 CRC64;

Query Match  
Best Local Similarity 2.6%; Score 16; DB 10; Length 275;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 482 FLDRTFESLHYSTMF 497  
Db 8 FLDRTFESLHYSTMF 23

Search completed: January 22, 2002, 16:36:44  
Job time: 344 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:26:20 ; Search time 22.77 Seconds  
(without alignments)  
615.703 Million cell updates/sec

Title: US-09-485-529-7

Perfect score: 623

Sequence: 1 MKREYDAGSGGGGGGMS.....TLGNHTRPLATSAWRLGP 623

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2.6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2.6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2.6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2.6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2.6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2.6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	3.0	259	4	US-09-117-853-6
2	19	3.0	282	4	US-09-117-853-8
3	19	3.0	532	4	US-09-117-853-2
4	14	2.2	221	4	US-09-117-853-4
5	12	1.9	475	2	US-08-861-464-14
6	12	1.9	475	2	US-08-396-001-14
7	12	1.9	475	4	US-09-323-433A-14
8	10	1.6	166	4	US-09-117-853-10
9	10	1.6	193	2	US-08-679-765-3
10	10	1.6	193	2	US-09-196-525-3
11	10	1.6	193	4	US-09-318-317-3
12	10	1.6	568	1	US-08-320-559-30
13	10	1.6	568	1	US-08-545-860D-30
14	10	1.6	568	5	PCR-US94-04496-30
15	10	1.6	590	2	US-08-785-310A-5
16	10	1.6	594	2	US-08-785-310A-6
17	9	1.4	11	1	US-08-785-310A-43
18	9	1.4	11	1	US-08-040-546-43
19	9	1.4	11	6	US-08-466-344-43
20	9	1.4	11	6	5206152-12
21	9	1.4	76	6	5496550-6
22	9	1.4	130	2	US-08-630-822A-90
23	9	1.4	130	2	US-09-005-068-90
24	9	1.4	148	1	US-08-207-904-15
25	9	1.4	180	6	5273901-7
26	9	1.4	180	6	5482709-6
27	9	1.4	216	3	US-08-506-553C-8
					US-08-506-553C-23

28	9	1.4	218	4	US-09-068-655-7	Sequence 7, Appl
29	9	1.4	246	4	US-09-185-160-7	Sequence 7, Appl
30	9	1.4	262	1	US-08-106-981-6	Sequence 6, Appl
31	9	1.4	276	3	US-08-506-553C-26	Sequence 26, Appl
32	9	1.4	285	3	US-08-482-085B-20	Sequence 20, Appl
33	9	1.4	285	4	US-09-318-661-4	Sequence 4, Appl
34	9	1.4	298	4	US-09-318-661-2	Sequence 2, Appl
35	9	1.4	300	3	US-08-765-856-2	Sequence 2, Appl
36	9	1.4	300	4	US-08-935-009A-2	Sequence 2, Appl
37	9	1.4	302	3	US-08-765-856-4	Sequence 4, Appl
38	9	1.4	302	4	US-08-935-009A-4	Sequence 4, Appl
39	9	1.4	310	3	US-08-651-136C-22	Sequence 22, Appl
40	9	1.4	322	1	US-08-014-943A-2	Sequence 2, Appl
41	9	1.4	322	1	US-08-486-421-3	Sequence 3, Appl
42	9	1.4	322	1	US-08-470-911-3	Sequence 3, Appl
43	9	1.4	322	2	US-08-486-809-3	Sequence 3, Appl
44	9	1.4	345	2	US-08-282-197C-50	Sequence 50, Appl
45	9	1.4	360	3	US-08-899-437-7	Sequence 7, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-117-853-6
; Sequence 6, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jintong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117,853
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: GB 9602796.6
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-117-853-6

Query Match 3.0%; Score 19; DB 4; Length 259;
Best local Similarity 100.0%; Pred. No. 4.4e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 CPTLKFHFTANOAILLEAF 332
DB 229 CPTLKFHFTANOAILLEAF 247

RESULT 2
US-09-117-853-8
; Sequence 8, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jintong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117,853
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: PCT/GB97/00390
; EARLIER FILING DATE: 1997-02-12
```

```

; EARLIER APPLICATION NUMBER: GB 9602796.6
; EARLIER FILING DATE: 1996-02-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-117-853-8

Query Match
3.0%; Score 19; DB 4; Length 282;
Best Local Similarity 100.0%; Pred. No. 4.8e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPLYKFAHFTANQALILEAF 332
|||||
Db 229 CPLYKFAHFTANQALILEAF 247

RESULT 3
US-09-117-853-2
; Sequence 2, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jinrong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117,853
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: PCT/GB97/00390
; EARLIER FILING DATE: 1997-02-12
; EARLIER APPLICATION NUMBER: GB 9602796.6
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-117-853-2

Query Match
3.0%; Score 19; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPLYKFAHFTANQALILEAF 332
|||||
Db 246 CPLYKFAHFTANQALILEAF 264

RESULT 4
US-09-117-853-4
; Sequence 4, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jinrong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117,853
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: PCT/GB97/00390
; EARLIER FILING DATE: 1997-02-12
; EARLIER APPLICATION NUMBER: GB 9602796.6
; NUMBER OF SEQ ID NOS: 12
```

```

; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-117-853-4

Query Match
2.2%; Score 14; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 233 RLVHALLACAEAVQ 246
|||||
Db 150 RLVHALLACAEAVQ 163

RESULT 5
US-08-861-464-14
; Sequence 14, Application US/08861464
; Patent No. 5874210
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence
; TITLE OF INVENTION: In Yeast
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/861,464
; FILING DATE: 22-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/396,001
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09351
; FILING DATE: 15-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,408
; FILING DATE: 16-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6408A22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 781-861-6240
; TELEFAX: 781-861-9540
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 475 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-861-464-14

Query Match
1.9%; Score 12; DB 2; Length 475;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 183 SSTSSSSSSSS 194  
 |||||||  
 DB 182 SSTSSSSSSSS 193

## RESULT 6

US-08-396-001-14  
 ; Sequence 14, Application US/08396001  
 ; Patent No. 5919618  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Guarente, Leonard P.  
 ; APPLICANT: Austriaco Jr., Nicanor  
 ; APPLICANT: Claus, James  
 ; APPLICANT: Cole, Francesca  
 ; APPLICANT: Kennedy, Brian  
 ; TITLE OF INVENTION: Genes Determining Cellular Senescence in  
 ; NUMBER OF SEQUENCES: 16  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 ; STREET: Two Militia Drive  
 ; CITY: Lexington  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02173  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/396,001  
 ; FILING DATE: 28-FEB-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Granahan, Patricia  
 ; REGISTRATION NUMBER: 32,227  
 ; REFERENCE/DOCKET NUMBER: MIT-6408A2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-861-6240  
 ; TELEFAX: 617-861-9940  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 475 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-396-001-14

Query Match 1.9%; Score 12; DB 2; Length 475;  
 Best Local Similarity 100.0%; Pred. No. 0.014;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 183 SSTSSSSSSSS 194  
 |||||||  
 DB 182 SSTSSSSSSSS 193

RESULT 7  
 US-09-323-433A-14  
 ; Sequence 14, Application US/09323433A  
 ; Patent No. 6218512  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Guarente, Leonard P.  
 ; APPLICANT: Austriaco Jr., Nicanor  
 ; APPLICANT: Claus, James J.  
 ; APPLICANT: Cole, Francesca  
 ; APPLICANT: Kennedy, Brian  
 ; TITLE OF INVENTION: Genes Determining Cellular Senescence in  
 ; NUMBER OF SEQUENCES: 5  
 ; FILE REFERENCE: 0050.1491-003  
 ; CURRENT APPLICATION NUMBER: US/09/323,433A

; CURRENT FILING DATE: 1999-06-01  
 ; PRIOR APPLICATION NUMBER: US 08/396,001  
 ; PRIOR FILING DATE: 1995-02-28  
 ; PRIOR APPLICATION NUMBER: PCT/US94/09351  
 ; PRIOR FILING DATE: 1994-08-15  
 ; PRIOR APPLICATION NUMBER: US 08/107,408  
 ; PRIOR FILING DATE: 1993-08-16  
 ; NUMBER OF SEQ ID NOS: 48  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 14  
 ; LENGTH: 475  
 ; TYPE: PRT  
 ; ORGANISM: Saccharomyces cerevisiae  
 ; US-09-323-433A-14

Query Match 1.9%; Score 12; DB 4; Length 475;  
 Best Local Similarity 100.0%; Pred. No. 0.014;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 183 SSTSSSSSSSS 194  
 |||||||  
 DB 182 SSTSSSSSSSS 193

RESULT 8  
 US-09-117-853-10  
 ; Sequence 10, Application US/09117853  
 ; Patent No. 6307126  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Harberd, Nicholas P  
 ; APPLICANT: Peng, Jintong  
 ; APPLICANT: Carol, Pierre  
 ; APPLICANT: Richards, Donald E  
 ; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana  
 ; FILE REFERENCE: 620-45  
 ; CURRENT APPLICATION NUMBER: US/09/117,853  
 ; CURRENT FILING DATE: 1998-08-12  
 ; EARLIER APPLICATION NUMBER: PCT/GB97/00390  
 ; EARLIER FILING DATE: 1997-02-12  
 ; EARLIER APPLICATION NUMBER: GB 9602796.6  
 ; EARLIER FILING DATE: 1996-02-12  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 10  
 ; LENGTH: 166  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 ; US-09-117-853-10

Query Match 1.6%; Score 10; DB 4; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 0.3;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 55 DVAKLEOLE 64  
 |||||||  
 DB 27 DVAKLEOLE 36

RESULT 9  
 US-08-679-765-3  
 ; Sequence 3, Application US/08679765  
 ; Patent No. 5840866  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Au-Young, Janice  
 ; APPLICANT: Goli, Surya K.  
 ; APPLICANT: Hillman, Jennifer L.  
 ; TITLE OF INVENTION: A NOVEL HUMAN UBQUITIN-CONJUGATING ENZYME  
 ; NUMBER OF SEQUENCES: 5  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive

CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/679,765  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0093 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 193 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1064914  
US-08-679-765-3

Query Match 1.6%; Score 10; DB 2; Length 193;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 STSSSSSSSS 193  
DB 9 STSSSSSSSS 18

RESULT 10  
US-09-196-525-3  
Sequence 3, Application US/09196525  
Patent No. 5989883  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Goli, Surya K.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING  
TITLE OF INVENTION: ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/196,525  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/679,765  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0093 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 193 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1064914  
US-09-196-525-3

Query Match 1.6%; Score 10; DB 2; Length 193;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 STSSSSSSSS 193  
DB 9 STSSSSSSSS 18

RESULT 11  
US-09-318-317-3  
Sequence 3, Application US/09318317  
Patent No. 6172199  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Goli, Surya K.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING  
TITLE OF INVENTION: ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/318,317  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,525  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0093 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 193 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1064914  
US-09-318-317-3

Query Match 1.6%; Score 10; DB 4; Length 193;

Best Local Similarity 100.0%; Pred. No. 0.35; Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 TSSSSSSSS 193

Db 9 STSSSSSSSS 18

RESULT 12

US-08-320-559-30

; Sequence 30, Application US/08320559

; Patent No. 5633135

; GENERAL INFORMATION:

; APPLICANT: Croce, Carlo

; APPLICANT: Canaanl, Eli

; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for

; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias

; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the

; TITLE OF INVENTION: All-1 Region

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135rfs

; STREET: One Liberty Place - 46th Floor

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: WORDPERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/320,559

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/062,443

; FILING DATE: 14 MAY 1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/971,094

; FILING DATE: 30-OCT-92

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/888,830

; FILING DATE: 27-MAY-92

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/805,093

; FILING DATE: 11-DEC-91

; ATTORNEY/AGENT INFORMATION:

; NAME: Deluca, Mark

; REGISTRATION NUMBER: 33,229

; REFERENCE/DOCKET NUMBER: TJU-0855

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (215) 568-3100

; TELEFAX: (215) 568-3439

; INFORMATION FOR SEQ ID NO: 30:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 568 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-320-559-30

Query Match 1.6%; Score 10; DB 1; Length 568;

Best Local Similarity 100.0%; Pred. No. 1; Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 185 TSSSSSSSS 194

Db 9 STSSSSSSSS 18

Db 148 TSSSSSSSS 157

RESULT 13

US-08-545-860D-30

; Sequence 30, Application US/08545860D

; Patent No. 6040140

; GENERAL INFORMATION:

; APPLICANT: Croce, Carlo

; APPLICANT: Canaanl, Eli

; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods

; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias

; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region

; NUMBER OF SEQUENCES: 94

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &

; ADDRESS: No. 6040140rfs

; STREET: One Liberty Place, 46th floor

; CITY: Philadelphia

; STATE: Pennsylvania

; COUNTRY: USA

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/545,860D

; FILING DATE: 07-MAR-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/04496

; FILING DATE: 22-APR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/10930

; FILING DATE: 09-DEC-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/327,392

; FILING DATE: 19-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/320,559

; FILING DATE: 11-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/062,443

; FILING DATE: 14-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/971,094

; FILING DATE: 30-OCT-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/888,839

; FILING DATE: 27-MAY-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/805,093

; FILING DATE: 11-DEC-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Deluca Esq., Mark

; REGISTRATION NUMBER: 33,229

; REFERENCE/DOCKET NUMBER: TJU-1262

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (215) 568-3100

; TELEFAX: (215) 568-3439

; INFORMATION FOR SEQ ID NO: 30:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 568 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-545-860D-30

Query Match 1.6%; Score 10; DB 3; Length 568;

Best Local Similarity 100.0%; Pred. No. 1;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 TSSSSSSSS 194  
|||||  
Db 148 TSSSSSSSS 157

## RESULT 14

PCT-US94-04496-30  
; Sequence 30, Application PC/TUS9404496  
; GENERAL INFORMATION:  
; APPLICANT: Croco, Carlo  
; APPLICANT: Canaan, Eli  
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &  
; ADDRESSEE: Norris  
; STREET: One Liberty Place, 46th floor  
; City: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/04496  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deluca Esq., Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: TJU-1242  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 568 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US94-04496-30

Query Match 1.6%; Score 10; DB 5; Length 568;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 TSSSSSSSS 194  
|||||  
Db 148 TSSSSSSSS 157

## RESULT 15

US-08-785-310A-5  
; Sequence 5, Application US/08785310A  
; Patent No. 5840532  
; GENERAL INFORMATION:  
; APPLICANT: McKnight, Steven L.  
; APPLICANT: Russell, David W.  
; TITLE OF INVENTION: Neuronal PAS Domain Protein  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA

COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/785,310A  
; FILING DATE: 21-JAN-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UTSD:1226  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 590 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-785-310A-5

Query Match 1.6%; Score 10; DB 2; Length 590;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 SSSSSSSSL 195  
|||||  
Db 214 SSSSSSSSL 223

Search completed: January 22, 2002, 16:30:53  
Job time: 273 sec



---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2002, 02:37:31 ; Search time 3020.32 Seconds  
(without alignments)

11606.885 Million cell updates/sec

Title: US-09-485-529-14

Sequence: 1 atagagagcgaggtagctc.....tggatggacgacgaactccg 2125

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_om: \*  
20: em\_or: \*  
21: em\_ov: \*  
22: em\_pat: \*  
23: em\_ph: \*  
24: em\_pl: \*  
25: em\_ro: \*  
26: em\_sts: \*  
27: em\_sy: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_htgo\_hum: \*  
31: em\_htgo\_inv: \*  
32: em\_htgo\_rod: \*  
33: em\_htg\_hum: \*  
34: em\_htg\_inv: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_other: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2125	100.0	2125	6	AX005805	AX005805 Sequence
2	1821	85.7	1872	6	TAEE242531	AJ242531 Triticum
3	1453	68.4	1768	6	AX005804	AX005804 Sequence
4	659	31.0	2709	6	AX005794	AX005794 Sequence
5	652	30.7	1746	6	AX005848	AX005848 Sequence
6	280	13.2	377	6	AX005862	AX005862 Sequence
7	251	11.8	453	6	AX005857	AX005857 Sequence
8	217	10.2	416	6	AX005809	AX005809 Sequence
9	210	9.9	425	6	AX005864	AX005864 Sequence
10	159	7.5	332	6	AX005849	AX005849 Sequence
11	157	7.4	324	6	AX005867	AX005867 Sequence
12	146	6.9	436	6	AX005863	AX005863 Sequence
13	122	5.7	511	6	AX005854	AX005854 Sequence
14	107	5.0	357	6	AX005853	AX005853 Sequence
15	94	4.4	230	6	AX005861	AX005861 Sequence
16	89	4.2	369	6	AX005852	AX005852 Sequence
17	81	3.8	211	6	AX005851	AX005851 Sequence
18	76	3.6	309	6	AX005855	AX005855 Sequence
19	65	3.1	1890	8	ZMA242530	AJ242530 Zea mays
20	65	3.1	2255	6	AX005806	AX005806 Sequence
21	57	2.7	285	6	AX005865	AX005865 Sequence
22	53	2.5	2500	8	AB030956	AB030956 Oryza sat
23	53	2.5	122497	8	AC087797	AC087797 Oryza sat
24	52	2.4	371	6	AX005808	AX005808 Sequence
25	52	2.4	377	6	AX005850	AX005850 Sequence
26	51	2.4	51	6	AX005896	AX005896 Sequence
27	48	2.3	259	6	AX005866	AX005866 Sequence
28	48	2.3	399	6	AX005856	AX005856 Sequence
29	45	2.1	800	8	AF377621	AF377621 Zea mays
30	45	2.1	800	8	AF377622	AF377622 Zea mays
31	45	2.1	800	8	AF377623	AF377623 Zea mays
32	45	2.1	800	8	AF377624	AF377624 Zea mays
33	45	2.1	800	8	AF377625	AF377625 Zea mays
34	45	2.1	800	8	AF377626	AF377626 Zea mays
35	45	2.1	800	8	AF377627	AF377627 Zea mays
36	45	2.1	800	8	AF377628	AF377628 Zea mays
37	45	2.1	800	8	AF377629	AF377629 Zea mays
38	45	2.1	800	8	AF377630	AF377630 Zea mays
39	45	2.1	800	8	AF377631	AF377631 Zea mays
40	45	2.1	800	8	AF377632	AF377632 Zea mays
41	45	2.1	800	8	AF377633	AF377633 Zea mays
42	45	2.1	800	8	AF377634	AF377634 Zea mays
43	45	2.1	800	8	AF377635	AF377635 Zea mays
44	45	2.1	800	8	AF377636	AF377636 Zea mays
45	45	2.1	800	8	AF377637	AF377637 Zea mays

## ALIGNMENTS

RESULT 1  
LOCUS AX005805 2125 bp DNA  
DEFINITION Sequence 14 from Patent WO9909174.  
ACCESSION AX005805  
VERSION AX005805.1 GI:9928802  
KEYWORDS  
SOURCE  
ORGANISM  
Triticum aestivum  
bread wheat.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
REFERENCE  
AUTHORS Harberd, N.P. and Peng, J.  
TITLE Genetic control of plant growth and development  
JOURNAL Patent: WO 9909174-A 14 25-FEB-1999;  
HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)  
FEATURES  
1..2125  
Location/Qualifiers  
/organism="Triticum aestivum"



Db	2041	GAATTGACGACCGCGCTACATGCATGGGCTACGGGTGGTCCGTCGCTGGCGTGA	2100
QY	2101	agaagtgatgacgacgaactccg	2125
Db	2101	AGAGGTGATGACGACGACTCG	2125
RESULT	2		
LOCUS	TAE242531		
DEFINITION	Triticum aestivum rht-D1a gene for gibberellin response modulator.		
ACCESSION	AJ242531		
VERSION	AJ242531.1		
KEYWORDS	gibberellin response modulator; rht-D1a gene.		
SOURCE			
ORGANISM	Triticum aestivum		
REFERENCE	1 (bases 1 to 1872)		
AUTHORS	Peng, J., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F., Sudhakar, D., Christou, P., Snape, J.W., Gale, M.D. and Harberd, N.P.		
TITLE	'Green revolution' genes encode mutant gibberellin response modulators		
JOURNAL	Nature 400 (6741), 256-261 (1999)		
MEDLINE	99347734		
REFERENCE	2 (bases 1 to 1872)		
AUTHORS	Harberd, N.P., Peng, J. and Richards, D.E.		
TITLE	Green revolution genes encode mutant gibberellin response modulators		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 1872)		
AUTHORS	Richards, D.E.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-MAY-1999) Richards D.E., Molecular Genetics, John Innes Centre, Colney Lane, NR4 7UJ, UNITED KINGDOM		
FEATURES			
SOURCE	1. 1872		
gene	/organism="Triticum aestivum"		
CDS	/db_xref="taxon:4565"		
	1. 1872		
	/gene="rht-D1a"		
	1. 1872		
	/gene="rht-D1a"		
	/note="GAI ortholog"		
	/codon_start=1		
	/product="gibberellin response modulator"		
	/protein_id="CAB5155.1"		
	/db_xref="GI:5640157"		
	/translatability="MKREYODAGSGGGGSGGSGSSEDKMVSAAAGBEEVDELLAALGYKRAISDMDVAOKLEOLEMAMGMGVGAACAAADSDPATLADTYVHYNPTDSSWWMESMSELNADSVPRPAPOLNATSSRYTSGSGGYFDLPVSDDSSITVLRPIIDSPAGAAATPAAVVDVDGAGIRLVHALCAEAAVQENLSAAELVKOIPLLASOGGAMKRVAAVYFGALAAVRFRFRPPDPSLIDAAFLDILAHRYESCPIKFAHFTANQILIAFVGCRRVHVVDGFIKQGMQMPALIDALPRGPPSRILGVPQPDDELDALQGVKMLAKOFATTVIRVDFOYRGLVAATLADLEPMLQPGEDPNEPEVIAVNSVFEMHRLIAOQGALEKVIQTVAVRVRPIVTVYVEOGAHHNSGFLDRFTESLHYTSMDSLEGSSGGGSPSEVSSGAAPAAAGTDQWSEYTLGRQITNVVACBQAEETKHEHTLGGWRNRLGNNGFEFVRLHGSNNYKQASTLLALFAGGDDGYKEEKGCULTLGMHTPLATLSAMRLAGP		
BASE COUNT	265 a 700 c 641 g 266 t		
ORIGIN			
QY	28	atgaagcgaggataccagcagcgcgagggcgagcgcgcgcgcggtgagcgcatgagctcg	87
Db	1	ATGAAGCGGGAATTACACGACCGCGAGGACCGCGCGCGGTGGCGCATGGAGCTCG	60
Query Match	85.7%	Score 1821;	DB 8; Length 1872;
Best Local Similarity	99.9%	Pred. No. 0;	
Matches 1871; Conservative	0;	Mismatches	1; Indels 0; Gaps 0;

[illegible]

QY 1168 accgagccctcgaagctggtggaagctcgcccaattctcgacacacatccgcgtc 1227  
|||||  
Db 1141 ACCGAGCCCTCGAGCAGAGTGGGCTGGAAGCTCGCCAGTTCGGCAGACACATCCGCTC 1200  
QY 1228 gacttcagttacgcgcgtctgtcgccgcaacgctcgagaccttgagacgcttcatgctg 1287  
|||||  
Db 1201 GACTTCGACTACCGCGCTCGTCTGCGCCGACGCTCGGACCTGGAGCGCTTCATGCTG 1260  
QY 1288 cagccggaagcggaagaaaccggaagcgagcccgagtaatcgccgtcaactagtc 1347  
|||||  
Db 1261 CAGCCGAGAGCGGAGAGAGACCCGAAAGAGAGCCGAGGTATATCCCTCACTATAGTC 1320  
QY 1348 ttcgagatgcaacgcgtctgtcgcgacgagcccgcccttgagaaaggttccttggcaccgtg 1407  
|||||  
Db 1321 TTCGAGATGACACCGCTGCTCGCGAGCCGCGGCTTGAGAAAGTCTCTGGGACACGCTG 1380  
QY 1408 cgcgcgcgtcgccgacaggtctgtcaacgctgtgtgagcagaagcgatatacaaatccgcgc 1467  
|||||  
Db 1381 CGCGCGTTCGGGCCCGAGATGCTCACCGGTGTGAGACAGAGCGGATACAACTCCGGC 1440  
QY 1468 aatttcctgagccgcttcaacgagctctgtcaactactatcaaccaagtctgattccctc 1527  
|||||  
Db 1441 ACATTCCTGAGACCGCTTCACCAAGTCTTGCACTACTTCCACATGTTCCATTCCCTC 1500  
QY 1528 gagggcgagctccgagcgcgcccaatccgaagtcaatcgagggtctgtctgctcct 1587  
|||||  
Db 1501 GAGGGGCGAGCTCGCGGCGGCCCATCCGAAGTCTCATCGGGGCTGCTGCTGCTCT 1560  
QY 1588 ggcgcgcgcgcgacgagcagagctatgtccgaggtgttaactcgccgcgagatctgcaac 1647  
|||||  
Db 1561 GCGCGCGCGGACGAGCAGGCTCATGTCCGAGGTGTACTCGCGCGGAGATCTGCAAC 1620  
QY 1648 gtgtgtcctcgagagggcgagcgagcagagcgacagagcgacagagcgcttgagcaatgagcg 1707  
|||||  
Db 1621 GGTGTGGCTTGAGAGGGGGGCGAGCGCAGAGCGCAGAGACGTGGGCCAGTGGCGG 1680  
QY 1708 aacgcgcttgagcaacgcggtctcgagacgctccacgttggtccaaatgcttacaagcag 1767  
|||||  
Db 1681 AACCGGCTTGAGACCGCGGTTCGAGACCGTTCACCTGGGCTCCAAATCCTTACAAGCAG 1740  
QY 1768 gagagcagcgtctgtgagctcttgccgagcgagcgagcttaagaagtgagagaagaga 1827  
|||||  
Db 1741 GCGAGACGCTCTGGCGCTTTCGCGCGGCGGAGCGGTCAAGGTGAGAGAGAGAA 1800  
QY 1828 ggcctcctgaacgctggtgtgagcagacgcgcgcgtgatgcacactcgacatgagcgctg 1887  
|||||  
Db 1801 GCGTGCCTGACCTGGGGTGGGCAACAAGCGCGCTGATGCGCACCTCGGCATGGCGCTG 1860  
QY 1888 gccgggagcgatga 1899  
|||||  
Db 1861 GCCGGGCGCTGA 1872

RESULT 3  
AX005804 1768 bp DNA PAT 24-AUG-2000  
LOCUS AX005804  
DEFINITION Sequence 13 from Patent WO9909174.  
ACCESSION AX005804  
VERSION AX005804.1 GI:9928801  
KEYWORDS  
SOURCE bread wheat.  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 1768)  
REFERENCE 1 (bases 1 to 1768)  
AUTHORS Harberd,N.P. and Peng,J.  
TITLE Genetic control of plant growth and development  
JOURNAL Patent: WO 9909174-A 13-25-FEB-1999;  
HARBERD NICHOLAS PAUL (GB); PENG JINKONG (GB)  
FEATURES  
source 1..1768  
/organism="Triticum aestivum"

BASE COUNT 368 a 595 c 527 g 278 t  
ORIGIN  
Query Match 68.4%; Score 1453; DB 6; Length 1768;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1503; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
/db\_xref="taxon:4565"  
QY 622 gccagagctctgtgtgtagagctgccccgcggttcgagccgagcgagcagccccc 681  
|||||  
Db 1 GCCAGAGCTGTGTGTGAGGCTGCTCCGCGGTGCGCGGCGGCGCAACCGAGCGCC 60  
QY 682 ggcgtcgagctcgctgtgttcgaacgacgagagccgagatccgctgtgtgacgcgtg 741  
|||||  
Db 61 GCGCTCCCGTGTGTGTGTGTCGACAGCAGAGAGCGGGATTCGCTGTGTGACGCGCTG 120  
QY 742 ctgagctgagcgagagcgctgacgagagaacctccgcgcgagagcgctgtgtgaa 801  
|||||  
Db 121 CTGGCGTGGCGGAGGCGCTGCAGCAGAGAACCTCTCCGCGGAGGCGCTGTGTGAAG 180  
QY 802 cagataacctgtctgtgagcgcttccagagcgagcgatgagtgcaaggtcgccgaattc 861  
|||||  
Db 181 CAGATACCTTCTGTGGCGCGCTCCAGAGGCGCGGATGCGAAGGTCCGCCCTACTTGC 240  
QY 862 ggcagagccctcgccgcgcgctcttcgcgttcgcgcgcgcgcgcgcgcgcgcgcgcgc 921  
|||||  
Db 241 GCGAGAGCCCTTGCGCGCGCGCTTCCTTCGCTTCGCGCGCACCGAGACGCTCTCTCTC 300  
QY 922 gagcgcccttcgcgcgcgcctctccagcgcgaattctacagttcctgcccctacaa 981  
|||||  
Db 301 GACGCGCGCTTGCGCGAGCTCTCCACGCGGCACTTCTACGAGTCTGCGCCCTACTCAAG 360  
QY 982 ttcgcgcacttaccgc 1041  
|||||  
Db 361 TTCGGGCACTTACCGGCACACAGGCGCATCTGAGAGCGCTTGCGGCGCTGCGCGCGG 420  
QY 1042 cagctctgcaattcgcaatcaagcagagagtgacgtgagccgcacattctcagagccctc 1101  
|||||  
Db 421 CAGGTGTGAGATTGGGCACTCAAGCAGGAGGATGCAGTGCGCGGCACTTCTCAGGCGCTC 480  
QY 1102 gcccctcgctccgcgcgcgcctccctcctgctcgactcaacgcgcgtcgcccccgaagcg 1161  
|||||  
Db 481 GCCCTCCGCTCCGCGCGCGCTCTCCCTGTTCCGCTCACCGGCGTGGCGCCCGCAGCGG 540  
QY 1162 gagagagccagagccctgagagaggttgagcttgaaagctcgcccaattcgagacacatc 1221  
|||||  
Db 541 GACGAGACCGAGCCCTTGACAGAGGTGGGCTGGAAGCTGCCAAGTTCGCGCACACATC 600  
QY 1222 cgcgtcgaattccagtaaccgcgcgcgtctgtcgccgaacgctcgagaccttgagacgctc 1281  
|||||  
Db 601 CGCGTGAATTCCAGTACCGCGGCGCTCGCGCGCACGCTGCGGACCTGGAGCGCTTC 660  
QY 1282 atgtctgacgcgagagcgaggaagaaaccgaaagagagccgagtaatcgcgctaac 1341  
|||||  
Db 661 ATGTCTGACGCCGAGAGGCGAGAGAGACCCGAAAGAGACCCGAGGTATTCCTCTAAC 720  
QY 1342 tcagtcttgagatatcacagcgctgtctgcgagagccgcgcgcgcgcgcgcgcgcgcgc 1401  
|||||  
Db 721 TCAGTCTTCGAGATGACCGGCTGCTCGCGAGCCCGGCGGCTTGAGAGGCTCTGGGCG 780  
QY 1402 accgtgcgcgcgcgttcgc 1461  
|||||  
Db 781 ACCGAGCGCGCGCTGCGCGCGCGAGATGCTGACGAGTGGAGACAGAGGCGAATCACAA 840  
QY 1462 tccggaacattcctggaacgcttccagagctctctgcaactactactcaacatgttgcgt 1521  
|||||  
Db 841 TCGGGACACTTCTGTGAGCGCTTCAACCGAGTCTGCACTACTTCAACCACTGTTCAAT 900  
QY 1522 tccctcgaagagcgagcttcgc 1581  
|||||  
Db 901 TCCCTGAGAGGCGGAGCTCCGCGCGGCGGCGCCATCCGAGTCTCATCGGGGCTCTGCT 960



[illegible]

RESULT	5				
LOCUS	AX005848	1746 bp	DNA	PAT	24-AUG-2000
DEFINITION	Sequence 57 from Patent WO9909174.				
ACCESSION	AX005848				
VERSION	AX005848.1	GI:9928843			
KEYWORDS					
SOURCE	bread wheat.				
ORGANISM	Triticum aestivum				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;				
	Poaceae; Triticaceae; Triticum.				
REFERENCE	1 (bases 1 to 1746)				
AUTHORS	Harberd,N.P., and Peng,J.				
TITLE	Genetic control of plant growth and development				
JOURNAL	Patent: WO 9909174-A 57 25-FEB-1999;				
	HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)				
FEATURES	Location/Qualifiers				
source	1..1746				
	/organism="Triticum aestivum"				
	/db_xref="taxon:4565"				
	74				
misc_feature	/note="n is any nucleotide"				
	680				
misc_feature	/note="n is any nucleotide"				
	786				
misc_feature	/note="n is any nucleotide"				
	972				
misc_feature	/note="n is any nucleotide"				
	1038				
misc_feature	/note="n is any nucleotide"				
	1130				
misc_feature	/note="n is any nucleotide"				
	1170				
misc_feature	/note="n is any nucleotide"				
	1212				
misc_feature	/note="n is any nucleotide"				
	590 c 495 g 278 t				
BASE COUNT	375 a				
ORIGIN	8 others				
Query Match	30.7%; Score 652; DB 6; Length 1746;				

Best Local Similarity 100.0%; Pred. No. 9.9e-291;  
Matches 652; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]





Db 200 AGCTACAGCCTCCACCTCTCCACCGTCACGGGAGCGGGGCTACTTCGATCTCCGC 259

Qy 449 cctccgtgactcctccagcagcatctacgcgctgagcgcatccctcccgagcg 508  
|||||  
Db 260 CCTCGTGGACTCCCTCCAGCAGCATCTACGGCGCTCGGCCCATCTCCCGCGGCG 319

Qy 509 cgaagcgagcgagcgagcctgtctcgagcactccgtgagcgatcccaagcgatgagc 568  
|||||  
Db 320 CGACGCGCCCGCGCGAGCTGTCGCGGAGTCGCGGAGATCCCAAGCGATGCGCAGTG 379

Qy 569 gggagagcagcagcctcgtcgtcatctctctcgtc 605  
|||||  
Db 380 GCGGAGCAGCAGCTGTCGTCATCTCTCTCTCTGTC 416

RESULT 9  
AX005864  
LOCUS AX005864 425 bp DNA PAT 24-AUG-2000  
DEFINITION Sequence 73 from Patent WO9909174.  
ACCESSION AX005864  
VERSION AX005864.1 GI:9928859  
KEYWORDS  
SOURCE bread wheat.  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poidea; Triticeae; Triticum.  
REFERENCE 1 (bases 1 to 425)  
AUTHORS Harberd, N.P. and Peng, J.  
TITLE Genetic control of plant growth and development  
JOURNAL Patent: WO 9909174-A 73 25-FEB-1999;  
HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)  
FEATURES  
source 1..425  
location/Qualifiers  
misc\_feature 3  
misc\_feature 11  
misc\_feature 196  
misc\_feature 274  
misc\_feature 369  
misc\_feature 370  
misc\_feature 371  
misc\_feature 379  
misc\_feature 394  
misc\_feature 402  
misc\_feature 414  
misc\_feature 423  
BASE COUNT 91 a 125 c 103 g 94 t 12 others  
ORIGIN

Query Match 9.98; Score 210; DB 6; Length 425;  
Best Local Similarity 99.68; Pred. No. 4.5e-86;  
Matches 260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1865 tcgcaactcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1924  
|||||  
Db 12 TCGCCACCTCGGCGATGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 71

Qy 1925 tacacatcgtgagcagtagagacaacacagcccgcgagcgagcgagcgagcgagcgag 1984

Db 72 TACACATGCTGTGAGCATGAGAGACACAGCCCGCGCGCGCGCGCGCGCGCGCGCG 131  
|||||

Qy 1985 cgcagcgacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2044  
|||||  
Db 132 CGCACGACGCGAGCGACTTGTAGAGAGAGAACTTAAGTATGTCTGACTGACGCTGAT 191

Qy 2045 tgcagcgacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2104  
|||||  
Db 192 TGCANAGACCGGCTACGATCGATCGGGCTACGGGTGTTCCGTCCTCGGCGTGAAG 251

Qy 2105 gtgagtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2125  
|||||  
Db 252 GTGATGAGCAGCAGCAACTCCG 272

RESULT 10  
AX005849  
LOCUS AX005849 332 bp DNA PAT 24-AUG-2000  
DEFINITION Sequence 58 from Patent WO9909174.  
ACCESSION AX005849  
VERSION AX005849.1 GI:9928844  
KEYWORDS  
SOURCE bread wheat.  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poidea; Triticeae; Triticum.  
REFERENCE 1 (bases 1 to 332)  
AUTHORS Harberd, N.P. and Peng, J.  
TITLE Genetic control of plant growth and development  
JOURNAL Patent: WO 9909174-A 58 25-FEB-1999;  
HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)  
FEATURES  
source 1..332  
location/Qualifiers  
misc\_feature 19  
misc\_feature 33  
misc\_feature 131  
BASE COUNT 37 a 150 c 84 g 58 t 3 others  
ORIGIN

Query Match 7.58; Score 159; DB 6; Length 332;  
Best Local Similarity 100.0%; Pred. No. 1.9e-62;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1004 aggcattcctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1063  
|||||  
Db 174 AGGCCATCTCGGAGGCGTTCGCGCGGCTCGCGCGGCTCGCACGTCCTCAGCTTCG 233

Qy 1064 agcagggagtagtgagcgagcagcagcagcagcagcagcagcagcagcagcagcag 1123  
|||||  
Db 234 AGCAGGAGATGAGTGGCGCGGAGCTTTCAGAGGCCCTGCGCCTCGCTCGCGGCG 293

Qy 1124 cctcgttcgcctcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1162  
|||||  
Db 294 CCTCGTCCGCTCACCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 332

RESULT 11  
AX005867/c  
LOCUS AX005867 324 bp DNA PAT 24-AUG-2000  
DEFINITION Sequence 76 from Patent WO9909174.  
ACCESSION AX005867  
VERSION AX005867.1 GI:9928862  
KEYWORDS  
SOURCE bread wheat.  
ORGANISM Triticum aestivum

REFERENCE	Embryophyta: Viridiplantae; Streptophyta: Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta, Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.		
AUTHORS	1 (bases 1 to 324)		
TITLE	Harberd, N.P. and Peng, J.		
JOURNAL	Genetic control of plant growth and development		
FEATURES	Patent: WO 9909174-A 76 25-FEB-1999.		
SOURCE	HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)		
source	Location/Qualifiers		
	1..324		
misc_feature	/organism="Triticum aestivum"		
	/db_xref="taxon:4565"		
misc_feature	158		
	/note="n is any nucleotide"		
misc_feature	161		
	/note="n is any nucleotide"		
misc_feature	217		
	/note="n is any nucleotide"		
BASE COUNT	39 a 141 c 85 g 56 t 3 others		
ORIGIN			
Query Match	7.4% Score 157; DB 6; Length 324;		
Best Local Similarity	100.0%; Pred. NO. 1.6e-61;		
Matches 157; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	144	gctgctgagcagctcgtcgtgatacaagtgtagcgccctccagacatgtagcgagcgttgccagaa	203
Db	157	gctctgtgcgcgcgtcggtgacaaaggctgcgcgcctccgacatgtagcgagcgttgccagaa	98
QY	204	gctgtagacagctcgtagatgagccatgtaggagtggtggcggtggcgccgagccgccccga	263
Db	97	gctgtagacagctcgtgagatgagccatgtaggagtggtggcggtggcgccgagccgccccga	38
QY	264	cgacagctgcgcacccacccctgcgcagcagaaacgtg	300
Db	37	cgacagctgcgcacccacccctgcgcagcagaaacgtg	1
RESULT 12			
AX005863			
LOCUS	AX005863 436 bp DNA PAT 24-AUG-2000		
DEFINITION	Sequence 72 from Patent WO9909174.		
ACCESSION	AX005863		
VERSION	AX005863.1 GI:9928858		
KEYWORDS			
SOURCE	bread wheat.		
ORGANISM	Triticum aestivum		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.		
REFERENCE	1 (bases 1 to 436)		
AUTHORS	Harberd, N.P. and Peng, J.		
TITLE	Genetic control of plant growth and development		
JOURNAL	Patent: WO 9909174-A 72 25-FEB-1999;		
FEATURES	HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)		
source	Location/Qualifiers		
	1..436		
misc_feature	/organism="Triticum aestivum"		
	/db_xref="taxon:4565"		
misc_feature	1		
	/note="n is any nucleotide"		
misc_feature	21		
	/note="n is any nucleotide"		
misc_feature	54		
	/note="n is any nucleotide"		
misc_feature	60		
	/note="n is any nucleotide"		
misc_feature	225		
	/note="n is any nucleotide"		
misc_feature	267		
	/note="n is any nucleotide"		
misc_feature	364		

misc_feature	385	/note="n is any nucleotide"		
misc_feature	394	/note="n is any nucleotide"		
misc_feature	404	/note="n is any nucleotide"		
misc_feature	412	/note="n is any nucleotide"		
misc_feature	415	/note="n is any nucleotide"		
misc_feature	416	/note="n is any nucleotide"		
misc_feature	425	/note="n is any nucleotide"		
misc_feature	426	/note="n is any nucleotide"		
misc_feature	427	/note="n is any nucleotide"		
misc_feature	433	/note="n is any nucleotide"		
misc_feature	436	/note="n is any nucleotide"		
BASE COUNT	58 a 153 c 141 g 66 t	18 others		
ORIGIN				
Query Match 6.9%; Score 146; DB 6; Length 436;				
Best Local Similarity 100.0%; Pred. No. 1.8e-56;				
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0				
Qy	454	gtgcactctctcagcagcaltcagcgtcgtcgcgccatccctccgcgcgcgcgcg 513		
Db	76	gtcgcactctctccacacacatctacggcgctggcggccgatccctcccgcgcgacg 135		
Qy	514	ggcgccggcgacactgtctcgccgcgactccgtgcgggagatcccaagcgatcgacatggcggg 573		
Db	136	ggcgccggcgacactgtctcgccgcgactccgtgcgggagatcccaagcgatcgacatggcggg 195		
Qy	574	agcgacacctgtctgcatctctctc 599		
Db	196	agcagcacctgtctgcatctctctc 221		
RESULT 13				
AX005854	511 bp	DNA		
LOCUS	AX005854	511 bp DNA		
DEFINITION	Sequence 63 from Patent WO909174.	PAT		
ACCESSION	AX005854	24-AUG-2000		
VERSION	AX005854.1			
KEYWORDS	GI.9928849			
SOURCE	bread wheat.			
ORGANISM	Triticum aestivum			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticaceae; Triticum.			
AUTHORS	1 (bases 1 to 511)			
TITLE	Harberd N.P. and Peng J.			
JOURNAL	Genetic control of plant growth and development			
FEATURES	Patent: WO 909174-A 63 25-FEB-1999;			
source	HARBERD NICHOLAS PAUL (GB); PENG JINONG (GB)			
	location/Qualifiers			
	1..511			
	/organism="Triticum aestivum"			
	/db_xref="taxon:4565"			
misc_feature	7	/note="n is any nucleotide"		
misc_feature	17	/note="n is any nucleotide"		
misc_feature	149	/note="n is any nucleotide"		
misc_feature	263	/note="n is any nucleotide"		
misc_feature	/note="n is any nucleotide"			

misc_feature	361	/note="n is any nucleotide"	
misc_feature	362	/note="n is any nucleotide"	
misc_feature	413	/note="n is any nucleotide"	
misc_feature	438	/note="n is any nucleotide"	
misc_feature	479	/note="n is any nucleotide"	
misc_feature	484	/note="n is any nucleotide"	
misc_feature	491	/note="n is any nucleotide"	
misc_feature	496	/note="n is any nucleotide"	
misc_feature	510	/note="n is any nucleotide"	
BASE COUNT	125	a	137
ORIGIN		c	117
		g	119
		t	13
		others	

Query Match	5.7%	Score 122;	DB 6;	Length 511;
Best Local Similarity	100.0%	Pred. No. 2.3e-45;		
Matches 122; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1883 gctctggccgggcccgtgatactctgcgagcttctgaacgctgttaagtaacacatcgtatgacatcgg 1942

Db 18 GCCTGGCCGGGCCGATCTCGCGAGTTTGAACGCTGAAGTACACATCTGTAGCAGCG 77

Qy	1943	aggacaacacagcccgcgcgcccgctctccgcggaacgcagcagcagcact	2002
Db	78	AGGACACACAGCCCCGGGGCCGGCCGGCTCCGGCGAACGCACGCACGCAGCGACT	137

QY	2003	tg	2004
Db	138	TG	139

RESULT 14	LOCUS	SEQUENCE DEFINITION	ACCESION	VERSION	GI:928848	PAT	24-AUG-2000
AX005853	AX005853	Sequence 62 from Patent WO909174.	AX005853	AX005853.1			

ORGANISM  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooidae; Triticeae; Triticum.  
REFERENCE  
1 (bases 1 to 357)  
Harberd, N.P. and Peng, J.  
Genetic control of plant growth and development  
Patent: WO 9909174-A 62 25-FEB-1999.  
AUTHORS  
HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)  
JOURNAL  
FEATURES  
Location/Qualifiers

misc_feature	6	/note="n is any nucleotide"
misc_feature	57	/note="n is any nucleotide"
misc_feature	177	/note="n is any nucleotide"
misc_feature	235	/note="n is any nucleotide"
misc_feature	238	/note="n is any nucleotide"
misc_feature	251	/note="n is any nucleotide"
misc_feature	286	/note="n is any nucleotide"

misc_feature	293	/note="n is any nucleotide"	
misc_feature	294	/note="n is any nucleotide"	
misc_feature	295	/note="n is any nucleotide"	
misc_feature	296	/note="n is any nucleotide"	
misc_feature	328	/note="n is any nucleotide"	
misc_feature	331	/note="n is any nucleotide"	
misc_feature	332	/note="n is any nucleotide"	
misc_feature	333	/note="n is any nucleotide"	
misc_feature	333	/note="n is any nucleotide"	
BASE COUNT	57 a	125 c	100 g
ORIGIN		60 t	15 others

Query Match 5.0%; Score 107; DB 6; Length 357;  
Best Local Similarity 100.0%; Pred. No. 2,1e-38;  
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1501 tactactccaccatgttctgaattcctctgaaagcgcgagaatccgcggcgccccatccgaa    1560
        |||||||
DB      58 TACTACTCCACCATGTTCGATTCCTCGAGGGCGGCAGCCTCGGGCGGCCCATCCGAA    117

```

OY	1561	gtctcatcgaggagctgctgtcctcctgcggccggcaacgaacca	160
Db	118	gtctcatcgaggagctgctgctccttcgccgccgcggcacggaacca	164

RESULT 15			
AX005861			
LOCUS	AX005861	230 bp	DNA
DEFINITION	Sequence 70 from Patent WO909174.		
ACCESSION	AX005861		
VERSION	AX005861.1	GI:9928856	

ORGANISM  
Trifolium aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooidae; Triticeae; Triticum.  
REFERENCE  
1 (bases 1 to 230)  
Harberd, N.P., and Peng, J.  
Genetic control of plant growth and development  
Patent: WO 9909174-A 70 25-FEB-1999;  
HARBERD NICHOLAS PAUL (GB); PENG JINONG (GB)  
FEATURES  
1  
330  
Location/Qualifiers

misc_feature	4	/organism="Triticum aestivum"
		/db_xref="taxon:4565"
misc_feature	14	/note="n is any nucleotide"
misc_feature	202	/note="n is any nucleotide"
misc_feature	210	/note="n is any nucleotide"
misc_feature	214	/note="n is any nucleotide"
BASE COUNT	31 a 90 c 76 g 28 t	5 others
ORIGIN		

Query Match	4.4%	Score 94	DB 6	Length 230
Best Local Similarity	100.0%	Pred. No.	2.4e-32	
Matches 94	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Qy	205	ctggagcagctcagatggccatgggagatggcgcgctggcgccgcccgcgac	264
Db	54	CTGGAGCAGCTCGAGATGGCCATGGGGATGGGGCGCGCTGGCGCCGCCGCCGAC	113
Qy	265	gacagcttcgcaaccacctcgccacggacacg	298
Db	114	GACAGCTTCGCGCACCCACCTCGCCACGACACCG	147

Search completed: January 23, 2002, 04:54:37  
job time: 8226 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2002, 02:47:47 ; Search time 197.61 Seconds  
(without alignments)  
9219.251 Million cell updates/sec

Title: US-09-485-529-14

Perfect score: 2125  
Sequence: 1 atagagagcgagtagctc.....tgatgtgacgacgactccg 2125

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N.Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2125	100.0	2125	20	AA36279	Wheat Rht clone 5a
2	1453	68.4	1768	20	AA36278	Wheat Rht clone C1
3	659	31.0	2709	20	AA36275	Composite DNA sequ
4	652	30.7	1746	20	AA36285	Consensus cDNA seq
5	280	13.2	377	20	AA36268	DNA sequence obtai
6	251	11.8	453	20	AA36263	DNA sequence obtai
7	217	10.2	416	20	AA36283	Partial sequence o
8	210	9.9	425	20	AA36270	DNA sequence obtai
9	159	7.5	332	20	AA36255	DNA sequence obtai
10	157	7.4	324	20	AA36273	DNA sequence obtai
11	146	6.9	436	20	AA36269	DNA sequence obtai

12	122	5.7	511	20	AA36260	DNA sequence obtai
13	107	5.0	357	20	AA36259	DNA sequence obtai
14	94	4.4	230	20	AA36267	DNA sequence obtai
15	89	4.2	369	20	AA36258	DNA sequence obtai
16	81	3.8	211	20	AA36257	DNA sequence obtai
17	76	3.6	309	20	AA36261	DNA sequence obtai
18	65	3.1	2255	20	AA36280	Maize 1a1 genomic
19	57	2.7	285	20	AA36277	DNA sequence obtai
20	52	2.4	371	20	AA36282	Partial sequence o
21	52	2.4	377	20	AA36256	DNA sequence obtai
22	51	2.4	51	20	AA36284	Oligonucleotide de
23	48	2.3	259	20	AA36272	DNA sequence obtai
24	48	2.3	399	20	AA36262	DNA sequence obtai
25	44	2.1	302	20	AA36281	Partial sequence o
26	40	1.9	200	20	AA36266	DNA sequence obtai
27	35	1.6	35	20	AA36203	Primer used for se
28	35	1.6	35	20	AA36204	Primer used for se
29	35	1.6	770	20	AA36277	Rice EST D39460 se
30	34	1.6	725	20	AA36276	Partial cDNA sequ
31	27	1.3	27	20	AA36244	Primer used for se
32	27	1.3	35	20	AA36202	Primer used for se
33	27	1.3	2151	18	AA36267	Maize ZCARECROW ZC
34	27	1.3	2151	21	AA36298	Maize ZCR gene par
35	27	1.3	3510	21	AA36299	Maize Scarecrow nu
36	25	1.2	25	20	AA36233	Primer used for se
37	25	1.2	25	20	AA36214	Primer used for se
38	25	1.2	25	20	AA36226	Primer used for se
39	25	1.2	25	20	AA36213	Primer used for se
40	24	1.1	24	20	AA36215	Primer used for se
41	23	1.1	23	20	AA36231	Primer used for se
42	23	1.1	23	20	AA36232	Primer used for se
43	23	1.1	23	20	AA36228	Primer used for se
44	23	1.1	23	20	AA36229	Primer used for se
45	23	1.1	23	20	AA36207	Primer used for se

#### ALIGNMENTS

RESULT 1	AA36279	AA36279 standard; DNA; 2125 BP.
ID	AA36279	
XX	AA36279;	
AC	16-JUL-1999	(first entry)
DT		
XX		
DE	Wheat Rht clone 5a1 genomic sequence.	
XX		
KW	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;	
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;	
KW	paclobutrazol; ss.	
XX		
OS	Triticum aestivum.	
XX		
PN	MO9909174-AL.	
XX		
PD	25-FEB-1999.	
XX		
PF	07-AUG-1998;	98WO-GB02383.
XX		
PR	13-AUG-1997;	97GB-0017192.
XX		
PA	(PLAN-) PLANT BIOSCIENCE LTD.	
XX		
PI	Harberd NP, Peng J, Richards DE;	
XX		
DR	WPI; 1999-181040/15.	
XX	P-PsDB; AAT02340.	
XX		
PT	New Triticum aestivum polynucleotides - encode a polypeptide which provides inhibition of the growth of plants, which inhibition is antagonised by gibberellin, used to confer a dwarf phenotype	

XX  
PS Disclosure; Fig 8a; 88pp; English.

The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in *Triticum Aestivum*, inhibit growth of the plant. This growth inhibition is antagonised by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds dwarf but let crop plants grow tall. The present sequence represents the wheat Rht clone 5a1 genomic sequence.

**SQ** Sequence 2125 BP; 325 A; 768 C; 723 G; 309 T; 0 other;

Query match	100.08; Score 2125; DB 20; Length 2125;
-------------	---

Matches	2125;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
---------	-------	--------------	----	------------	----	--------	----	------	----

QY 1 atagagagcgagtagctcgcgcatcatgaagcgggagttaccagagccggaggagc 60  
|||||  
Db 1 atagaagcgagtagctcgcgcatcatgaagcgggagttaccagagccggaggagc 60

	61	ggcgcgcgccgtctgcgcgcacatggctctcgtcccgaggaacaagtatgcttcgcgcgcgcgcg	12
QY			
D8	61	aggccccaccattaaccccataacctcatcccaagaagaatatataatcttccaaccccacca	12

OY 121 gggaagggggaagtgcacagctgtcgcgcgcctcgatcaaaagtgcgcgcctcc 18  
 |||  
 db 121 ggggaagggggaagtgcacagctgtcgcgcgcctcgatcaaaagtgcgcgcctcc 18

181 gacatggcgagcgtgcygcagaaactgagcagctcgatgcatgtggtggcgcgc 24

**QY**     **241** gtggcgccgcgcccgcacagcttcgccaccactcgcacgacaacctg 30  
|||||

301 cactacaaccacccgaactgtctctcttggctcgagacatgtctgtcgagctcaacgcg 36

DV CACCTCAACAACCCCAACGAGCAGCGCCGCGCCGCGAGCTCAACGCTTCAACTCTCTTCACCGTCAAG 42

DY .361 CCGCGCGCGCGCCCTCCCGGCCGCGCCGCGCCGCGAGCTCAACGCTTCAACTCTCTTCACCGTCAAG 42

D6 361 ccgcgcgcgcgccctccgcgcgcgcgcgcgaactcaagcgcctccaactctcttccacgcgtcacg 42

OY 421 ggcacgcgcgcgcctactcgatctccgcgcgccttcgctgcactctctccagcagcatctacgcg 48

**Dd** 421 ggcagcgcggtactatctgatctccgcccttcgtcgaacctctcaagaacattaacgcg 48  
**Gy** 481 ctgcgcgcgatctccctcccgcgcgcgcgaacgcgcgcgcgcgaactgtccgcgaactcc 54

Oy	541	qtgcgqgatalcccaagcgaatlgcgacactggcgaaggacagacactctgtcatctctctcc	60
Db	481	ctgcgagcgatcccctccccgccggcgagcgcgcgcgcgcgaacctgtccgcgaattcc	54

Dc 541 gtgcgggatacccaagcgatgcgcactggcggagcagcacctcgtcgtcatctctccc 60

**Dd**

601 |tcgtcgtctctcgtggtgggcgcagcgagtctctgtgtgtgtagtgccccgcgcgtcgcg| 66

D0  
661 gacgcgagccaaagcgaaccggcgctgccaggctgtgttgcaaacagcaggaagcgccggg 72

QY 721 attcgcgtgtgcacgcgctgtcgtgcctgcgcgagagccgtgcagcagagaaacctctc 780  
|||||

QY 781 gccgcggaagcgcctgctgaagcagataccctctgctgcccgcgtccacaggcggcgcgcatg 840

841 cgcacagctcgcgcctactctcgcgagagccctcgcgcgcgcgcgtcttcgcgccttcgcgcgc 900

DB 841 ccgcaaggtcgcgcctacttcgcgcgagggccctcgcgcgcgcgcgtcttcgcgccttcgcgcgcg 900

QY 901 cagccgcgacagctccctcctcgcagcgccgccttcgcgcgaactccttcacagcgcaactctac 960

Db 901 cagccggaagctccctcctctcgacgcgcgcctctgcgacctccctccacgcgcgaactctac 960

QY 961 gagtcctgcacctacctaagttcgcgcgaacttaacgcgcaacacgacatccttgagcgc 102

Db 961 gattctctgcacctcaagtctgcgcacttcaacgcgcaaccagccatctctgagggcg 102

Qy 1021 ttctgcgcgcctgcgcgcgcgcgtctcaacgtctgcacttcgcatcaagcagggatgcagtgc 108

0v	1081	ccgcacatttcacagacccctcagccctccatcccaagcgcctccatcatttcgcctcacc	114
Db	1021	ttcgcgcgcgtgcgcgcgtgcagctcgtcgaacttcgcatacgaagggaatgcagttg	108

[illegible]

Db 1141 ggcgtcggcccccgagccgcgagacgagacgcctcgagcaggtgggtggaagctc 120

Db 1201 gccacgttcgcgcacacacatccgcgttcgaacttcagtaaccgcgcgtctcgcgcgcacgc 126

Db 1261 ctcgcgacctgagccgtcactcagccgagggcgaggaacccgaacgagag 132

Db 1321 cccgagtgatcgcgtcaactcagctctcgagatgcacccgcgtctcgccagcccgcc 138

Ddb 1381 ggcctcgagaaagtccctcgtgcaaccgttcgtcgtccacgaatcgttcaacctcgtcgt 144

QY 1441 gagcaggaagcgaaatcaacaatccgcgacattcttgaccgcgttcaacgggtctctcgac 150  
|||||  
Db 1441 gagcaggaagcgaaatcaacaatccgcgacattcttgaccgcgttcaacgggtctctcgac 150

Qy 1501 tactactccacatgttcgattccctcgagggcggaagctccggcgcgccatccgaa 156  
|||||  
Db 1501 tactactccacatgttcgattccctcgagggcggaagctccggcgcgccatccgaa 156

DY 1561 gtccatcggggtctgtctgcctccgcgcgcgcgcacagaccagatcatgttccga 162  
|||||  
Dh 1561 attcatcaaaactactactactctcttcaccccacacacacacacatatccaa 162

QY 1621 gtgtacctcgccgcgcagatctgcaacgtgtgtgcctgcgcaggggcgcgcgcgcacagag 168

1681 ccgccagagacgcctggccagctggcggaaaccggcctgggcaacgcccgggttcgagacgcgc 174

QY 1741 cactgggtccatgctctacagcagcgacgcctgctgcgctcttcgcgcgcgc 180

D0 1741 caccctgggctcccaatctgcacaaagcagtcgaagacacgtcacctcggtggtccacccgctggc 180

DY 1801 gaacgctacaagaatcgaggaaggaaagacgtctccttgacctcggttggcacaacagctgccg 186



Db	1801	gacggctaccagaatggaagaaaggcgcgtctgaagctcggggttggcaacacgcgccg	1860
QY	1861	ctgatccgcaccctcggcatgtagcgccttgcccggcgcgtatcttcgcgaattttgaacctg	1920
Db	1861	ctgatccgcaccctcggcatgtagcgccttgcccggcgcgtatcttcgcgaattttgaacctg	1920
QY	1921	taagtacaacatcgtgaagcatgtaagaaacaacacagcccccgcgcccccgccttcocg	1980
Db	1921	taagtacaacatcgtgaagcatgtaagaaacaacacagcccccgcgcccccgccttcocg	1980
QY	1981	cgaacgcacgcacgcacgcacgcactttaagaagaagaagcctaattgtcatgtcaagtacgct	2040
Db	1981	cgaacgcacgcacgcacgcacgcactttaagaagaagaagcctaattgtcatgtcaagtacgct	2040
QY	2041	gaatttcgacgaccggcgcgtacgatactgggctacggtgtgttccttcgcgtctggcttga	2100
Db	2041	gaatttcgacgaccggcgcgtacgatactgggctacggtgtgttccttcgcgtctggcttga	2100
QY	2101	agaagttgatgtgacgcagcaactccg	2125
Db	2101	agaagttgatgtgacgcagcaactccg	2125
RESULT	2		
ID	AAX36278		
AC	AAX36278 standard; cDNA; 1768 BP.		
XX			
XX	AAX36278;		
DT	16-JUL-1999 (first entry)		
XX			
DE	Wheat Rht clone C15-1 cDNA sequence.		
XX			
KM	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;		
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;		
KW	pachlobutrazol; ss.		
OS	Triticum aestivum.		
XX			
FN	M09909174-A1.		
XX			
PD	25-FEB-1999.		
XX			
PF	07-AUG-1998; 98MO-G802383.		
XX			
PR	13-AUG-1997; 97GB-0017192.		
XX			
PA	(PLAN-) PLANT BIOSCIENCE LTD.		
XX			
PI	Harberd NP, Peng J, Richards DE;		
XX			
DR	WPI: 1999-181040/15.		
RR	P-PADB; AAY02539.		
PT	New Triticum Aestivum polynucleotides - encode a polypeptide which		
PT	provides inhibition of the growth of plants, which inhibition is		
PT	antagonised by gibberellin, used to confer a dwarf phenotype		
XX			
PS	Disclosure; Fig 7a; 88pp; English.		
XX			
CC	The specification describes polypeptides encoded by the Rht gene (and		
CC	its homologues) that, when expressed in Triticum Aestivum, inhibit		
CC	growth of the plant. This growth inhibition is antagonised by		
CC	gibberellin. The products can be used to provide Rht expression in		
CC	plants, conferring a dwarf phenotype on a plant which is correctable		
CC	by treatment with gibberellin. In addition, the products can be		
CC	used to produce Rht mutant plants which are dwarfed compared with		
CC	wild-type, the dwarfing being gibberellin-insensitive. Taller plants		
CC	may be made by knocking out Rht or the relevant homologous gene in		
CC	the plant of interest. Plants may be made which are resistant to		
CC	compounds which inhibit gibberellin biosynthesis, such as paclobutrazol		
CC	e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds		

[illegible]

[illegible]

PT provides inhibition of the growth of plants, which inhibition is  
PR antagonised by gibberellin, used to confer a dwarf phenotype  
XX  
PS Disclosure, Flg 3a, 86pp: English.

The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in *Triticum Aestivum*, inhibit growth of the plant. This growth inhibition is antagonised by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds dwarf but let crop plants grow tall. The present sequence represents the composite DNA sequence of wheat Rht gene.

Sequence 2709 BP; 473 A; 900 C; 823 G; 482 T; 31 other;

Query Match	31.0%;	Score 659;	DB 20;	Length 2709;
Best Local Similarity	100.0%;	Pred. No. 4.2e-274;		
Matches 659; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

AA36285	
ID	AA36285 standard; cDNA: 1746 BP.
XX	
AC	AA36285;
XX	
DT	16-JUL-1999 (first entry)
DE	
Consensus	cDNA sequence of wheat Rht clone C15-1.
KW	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
KM	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW	pachlobutrazol; ss.
OS	Triticum aestivum.
PN	MO9909174-A1.
PD	25-FEB-1999.
PF	07-AUG-1998; 98WO-GB02383.
PR	13-AUG-1997; 97GB-0017192.
PA	(PLAN-) PLANT BIOSCIENCE LTD.
PI	Harberd NP, Peng J, Richards DE;
DR	WPI: 1999-181040/15.
PT	New Triticum Aestivum polynucleotides - encode a polypeptide which
FT	provides inhibition of the growth of plants, which inhibition is
PS	antagonised by gibberellin, used to confer a dwarf phenotype
XX	
PS	Disclosure: Fig 2a; 88pp; English.
XX	
CC	The specification describes polypeptides encoded by the Rht gene (and
CC	its homologues) that, when expressed in Triticum Aestivum, inhibit
CC	growth of the plant. This growth inhibition is antagonised by
CC	gibberellin. The products can be used to provide Rht expression in
CC	plants, conferring a dwarf phenotype on a plant which is correctable
CC	by treatment with gibberellin. In addition, the products can be
CC	used to produce Rht mutant plants which are dwarfed compared with
CC	wild-type, the dwarfing being gibberellin-insensitive. Taller plants
CC	may be made by knocking out Rht or the relevant homologous gene in
CC	the plant of interest. Plants may be made which are resistant to
CC	compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
CC	e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
CC	dwarf but let crop plants grow tall. The present sequence
CC	represents the consensus cDNA sequence of wheat Rht clone C15-1.
XX	
SQ	Sequence 1746 BP; 375 A; 590 C; 495 G; 278 T; 8 other;
Query Match	30.7%; Score 652; DB 20; Length 1746;
Best Local Similarity	100.0%; Pred. No. 4.5e-271;
Matches 652; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	653 cgatcgccgcggcagcaacgcagaccgccgcgtctgcgtctgttgcacacgcaag 712 
Db	7 cgctcgccgcggcagcaacgcagaccgccgcgtctgcgtctgttgcacacgcaag 66 
OY	713 aggcgcggatttcgcttggtgtcacgcgcgtctgcgtctgcgcggaggcgtgtcacagaaga 772 
Db	67 aggcgcggatttcgcttggtgtcacgcgcgtctgcgtctgcgcggaggcgtgtcacagaaga 126 
OY	773 accttcgcgcgcggaggcgcgttgtgaagacgatcccttgccttgccgcgtccagggccg 832 
Db	127 accttcgcgcgcggaggcgcgttgtgaagacgatcccttgccttgccgcgtccagggccg 186 
OY	833 ggcgcgatgcgaaggtgcgcgcctacttcggcgaaggccctgcgccgcgcgtcttcgcgt 892 
Db	187 ggcgcgatgcgaaggtgcgcgcctacttcggcgaaggccctgcgccgcgcgtcttcgcgt 246 

OY	893	tccgcccgaagcgcggacagctccctctcgacgagcgcgttcgcgcgaactctccaagcgcg	952
Db	247	tcgcgccgcaagcggaaagctcctctctcgacgagcgcgttcgcgcgaactctccaagcgcg	306
OY	953	accttaagaatctctgccccctaactcaagtctgcgcgaattcacaccgaacaggcatcc	101
Db	307	aactttagaattctctgccccctaactcaagtctgcgcgaattcacaccgaacaggcatcc	366
OY	1013	tggagagcgtttccgcggcgtccgcgcgcgtgcacaagctgcagattccgatataagcaaggga	107
Db	367	tggagagcgtttccgcggcgtccgcgcgcgtgcacaagctgcagattccgatataagcaaggga	426
OY	1073	tgcagtgcccccaactcttcacagagcctctcgccctcgtctccgcgcgcctccctcttc	113
Db	427	tgcagtgcccccaactcttcacagagcctctcgccctcgtctccgcgcgcctccctcttc	486
OY	1133	gctcaaccgagctgcggcccccgcagccgagcagagacagagccctgacaagctggct	119
Db	487	gctcaaccgagctgcggcccccgcagccgagcagagacagagccctgacaagctggct	546
OY	1193	ggaagctgcgccagtltcggcgacacacatccgctgcgcagcttcacagtcggcctcgtc	125
Db	547	ggaagctgcgccagtltcggcgacacacatccgctgcgcagcttcacagtcggcctcgtc	606
OY	1253	cgcgcaagctgcgcgagcctgtgagagcgttatcgtctgcagccgagagcgagga	1304
Db	607	cgcgcaagctgcgcgagcctgtgagagcgttatcgtctgcagccgagagcgagga	658
 RESULT 5 AAK36268 ID AAK36268 standard; DNA; 377 BP.			
AC	AAK36268;		
DT	16-JUL-1999	(first entry)	
DE	DNA sequence obtained after sequencing wheat Rht clone 5a1.		
KM	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;		
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;		
XX	pachlobutrazol; ss.		
OS	Triticum aestivum.		
PM	W09909174-A1.		
PD	25-FEB-1999.		
Pf	07-AUG-1998;	98WO-GB02383.	
Pf	13-AUG-1997;	97GB-0017192.	
PA	(PLAN-) PLANT BIOSCIENCE LTD.		
PI	Harberd NP, Peng J, Richards DE;		
DR	WPI; 1999-181040/15.		
XX			
XX	The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in Triticum Aestivum, inhibit growth of the plant. This growth inhibition is antagonised by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants		
PT	antagonised by gibberellin, used to confer a dwarf phenotype		
PT			
PS	Disclosure; Fig 2c(1); 88pp; English.		



DR WPI; 1999-181040/15.  
 DR P-PSDB; AAY02544.  
 XX  
 PT New Triticum aestivum polynucleotides - encode a polypeptide which  
 PT provides inhibition of the growth of plants, which inhibition is  
 XX antagonised by gibberellin, used to confer a dwarf phenotype  
 PS  
 XX Disclosure; Fig 12a; 88pp; English.  
 CC The specification describes polypeptides encoded by the Rht gene (and  
 CC its homologues) that, when expressed in Triticum aestivum, inhibit  
 CC growth of the plant. This growth inhibition is antagonised by  
 CC gibberellin. The products can be used to provide Rht expression in  
 CC plants, conferring a dwarf phenotype on a plant which is correctable  
 CC by treatment with gibberellin. In addition, the products can be  
 CC used to produce Rht mutant plants which are dwarfed compared with  
 CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
 CC may be made by knocking out Rht or the relevant homologue gene in  
 CC the plant of interest. Plants may be made which are resistant to  
 CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
 CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
 CC dwarf but let crop plants grow tall. The present sequence represents  
 CC the partial sequence of the wheat Rht-10 allele.  
 CC  
 XX  
 SQ Sequence 416 BP; 59 A; 168 C; 129 G; 60 T; 0 other;  
 Query Match 10.2%; Score 217; DB 20; Length 416;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-84;  
 Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 389 agtcaacgctcaccctcctcaccgctcagcgagcgagcgctcttgcattccgcg 448  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 200 agtcaacgctcaccctcctcaccgctcagcgagcgagcgctcttgcattccgcg 259  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 449 cctcgcctgcgactcctcagcagcatctacgcgctgcgcgcgacccctccgcgcgcg 508  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 260 cctcgcctgcgactcctcagcagcatctacgcgctgcgcgcgacccctccgcgcgcg 319  
 OY 509 cgaagcgagcgagcgagcgtctccgcgactcgtgcggagatcccaagcgagtcgcgactg 568  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 320 cgaagcgagcgagcgagcgtctccgcgactcgtgcggagatcccaagcgagtcgcgactg 379  
 OY 569 gcggagcgagcagcctcgtctcattcctcctccctccgcgc 605  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 380 gcggagcgagcagcctcgtctcattcctcctccctccgcgc 416  
 RESULT 8  
 AAX36270  
 ID AAX36270 standard; DNA; 425 BP.  
 XX  
 AC AAX36270;  
 XX  
 DT 16-JUL-1999 (first entry)  
 XX  
 DE DNA sequence obtained after sequencing wheat Rht clone 5a1.  
 XX  
 KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
 KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
 KW paclobutrazol; ss.  
 XX  
 OS Triticum aestivum.  
 XX  
 PN WO9909174-A1.  
 XX  
 PD 25-FEB-1999.  
 XX  
 PF 07-AUG-1998; 98WO-GB02383.  
 XX  
 PR 13-AUG-1997; 97GB-0017192.  
 XX  
 PA (PLAN-) PLANT BIOSCIENCE LTD.

XX  
 PI Harberd NP, Peng J, Richards DE;  
 XX  
 DR WPI; 1999-181040/15.  
 XX  
 PT New Triticum aestivum polynucleotides - encode a polypeptide which  
 PT provides inhibition of the growth of plants, which inhibition is  
 XX antagonised by gibberellin, used to confer a dwarf phenotype  
 PS  
 XX Disclosure; Fig 2c(3); 88pp; English.  
 CC The specification describes polypeptides encoded by the Rht gene (and  
 CC its homologues) that, when expressed in Triticum aestivum, inhibit  
 CC growth of the plant. This growth inhibition is antagonised by  
 CC gibberellin. The products can be used to provide Rht expression in  
 CC plants, conferring a dwarf phenotype on a plant which is correctable  
 CC by treatment with gibberellin. In addition, the products can be  
 CC used to produce Rht mutant plants which are dwarfed compared with  
 CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
 CC may be made by knocking out Rht or the relevant homologue gene in  
 CC the plant of interest. Plants may be made which are resistant to  
 CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
 CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
 CC dwarf but let crop plants grow tall. The present sequence was  
 CC obtained after partially sequencing wheat Rht clone 5a1.  
 CC  
 XX  
 SQ Sequence 425 BP; 91 A; 125 C; 103 G; 94 T; 12 other;  
 Query Match 9.9%; Score 210; DB 20; Length 425;  
 Best Local Similarity 99.6%; Pred. No. 6.9e-81;  
 Matches 260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1865 tcgcaacctggagatggcgcttgcgcgcgcgctgactcgcgagtttgaacgtgtaag 1924  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 12 tcgcaacctggagatggcgcttgcgcgcgcgctgactcgcgagtttgaacgtgtaag 71  
 OY 1925 taacacatggaacatggaagaacacacagcccgagcgcccgccgctccgcgga 1984  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 72 taacacatggaacatggaagaacacacagcccgagcgcccgccgctccgcgga 131  
 OY 1985 cgaagcgagcgagcacttgaagaagaagctaaatgcatgctagtgagcgctgaat 2044  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 132 cgaagcgagcgagcacttgaagaagaagctaaatgcatgctagtgagcgctgaat 191  
 OY 2045 tgcagcgagcgagcactgagatcgagctgagctggttccgctcgtctgaggaag 2104  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 192 tgcagcgagcgagcactgagatcgagctgagctggttccgctcgtctgaggaag 251  
 OY 2105 gtgagtgagcagcgaactcgcg 2125  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 252 gtgagtgagcagcgaactcgcg 272  
 RESULT 9  
 AAX36255  
 ID AAX36255 standard; DNA; 332 BP.  
 XX  
 AC AAX36255;  
 XX  
 DT 16-JUL-1999 (first entry)  
 XX  
 DE DNA sequence obtained after sequencing wheat Rht clone 14a1.  
 XX  
 KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
 KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
 KW paclobutrazol; ss.  
 XX  
 OS Triticum aestivum.  
 XX  
 PN WO9909174-A1.  
 XX  
 PD 25-FEB-1999.



```

XX (PLAN-) PLANT BIOSCIENCE LTD.
XX
XX Harberd NP, Peng J, Richards DE;
XX
XX WPI; 1999-181040/15.
XX
XX New Triticum aestivum polynucleotides - encode a polypeptide which
XX PT provides inhibition of the growth of plants, which inhibition is
XX PT antagonised by gibberellin, used to confer a dwarf phenotype
XX
XX PS Disclosure; Fig 2c(2); 88pp; English.
XX
XX The specification describes polypeptides encoded by the Rht gene (and
XX CC its homologues) that, when expressed in Triticum aestivum, inhibit
XX CC growth of the plant. This growth inhibition is antagonised by
XX CC gibberellin. The products can be used to provide Rht expression in
XX CC plants, conferring a dwarf phenotype on a plant which is correctable
XX CC by treatment with gibberellin. In addition, the products can be
XX CC used to produce Rht mutant plants which are dwarfed compared with
XX CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants
XX CC may be made by knocking out Rht or the relevant homologous gene in
XX CC the plant of interest. Plants may be made which are resistant to
XX CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
XX CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
XX CC dwarf but let crop plants grow tall. The present sequence was
XX CC obtained after partially sequencing wheat Rht clone 5a1.
XX
XX SQ Sequence 436 BP; 58 A; 153 C; 141 G; 66 T; 18 other;

Query Match          6.9%; Score 146; DB 20; Length 436;
Best Local Similarity 100.0%; Pred. No. 2.3e-53;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 gtcgactctccagcagcatctacgctgcgagccgaccccccgcggcgcgagc 513
    |||||||
DB 76 gtcgactctccagcagcatctacgctgcgagccgaccccccgcggcgcgagc 135

QY 514 ggcgcggcgagcctgtccgcgcgactccgtgcgggattcccaagcgatgcgactggcggg 573
    |||||||
DB 136 ggcgcggcgagcctgtccgcgcgactccgtgcgggattcccaagcgatgcgactggcggg 195

QY 574 agcagcactcgtcgtcatctctctc 599
    |||||||
DB 196 agcagcactcgtcgtcatctctc 221

RESULT 12
AA36260
ID AAX36260 standard; DNA; 511 BP.
XX
XX AAX36260;
XX
XX 16-JUL-1999 (first entry)
XX
XX DNA sequence obtained after sequencing wheat Rht clone 14a1.
XX
XX Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
XX KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
XX KM paclobutrazol; ss.
XX
XX OS Triticum aestivum.
XX
XX PN WO9909174-A1.
XX
XX 25-FEB-1999.
XX
XX 07-AUG-1998; 98WO-GB02383.
XX
XX 13-AUG-1997; 97GB-0017192.
XX
XX (PLAN-) PLANT BIOSCIENCE LTD.

```

```

XX Harberd NP, Peng J, Richards DE;
XX
XX WPI; 1999-181040/15.
XX
XX New Triticum aestivum polynucleotides - encode a polypeptide which
XX PT provides inhibition of the growth of plants, which inhibition is
XX PT antagonised by gibberellin, used to confer a dwarf phenotype
XX
XX PS Disclosure; Fig 2b(6); 88pp; English.
XX
XX The specification describes polypeptides encoded by the Rht gene (and
XX CC its homologues) that, when expressed in Triticum aestivum, inhibit
XX CC growth of the plant. This growth inhibition is antagonised by
XX CC gibberellin. The products can be used to provide Rht expression in
XX CC plants, conferring a dwarf phenotype on a plant which is correctable
XX CC by treatment with gibberellin. In addition, the products can be
XX CC used to produce Rht mutant plants which are dwarfed compared with
XX CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants
XX CC may be made by knocking out Rht or the relevant homologous gene in
XX CC the plant of interest. Plants may be made which are resistant to
XX CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
XX CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
XX CC dwarf but let crop plants grow tall. The present sequence was
XX CC obtained after partially sequencing wheat Rht clone 14a1.
XX
XX SQ Sequence 511 BP; 125 A; 137 C; 117 G; 119 T; 13 other;

Query Match          5.7%; Score 122; DB 20; Length 511;
Best Local Similarity 100.0%; Pred. No. 4.8e-43;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1883 gcttgccgggcccgtgattctcgcgagtttgaacgctgtgaacatcgtgagcatgg 1942
    |||||||
DB 18 gcttgccgggcccgtgattctcgcgagtttgaacgctgtgaacatcgtgagcatgg 77

QY 1943 agaacacacagcccgcgccgcccgcgctctccggcgacgacgacgacgacact 2002
    |||||||
DB 78 agaacacacagcccgcgccgcccgcgctctccggcgacgacgacgacgacgacact 137

QY 2003 tg 2004
    ||
DB 138 tg 139

RESULT 13
AA36259
ID AAX36259 standard; DNA; 357 BP.
XX
XX AAX36259;
XX
XX 16-JUL-1999 (first entry)
XX
XX DNA sequence obtained after sequencing wheat Rht clone 14a1.
XX
XX Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
XX KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
XX KM paclobutrazol; ss.
XX
XX OS Triticum aestivum.
XX
XX PN WO9909174-A1.
XX
XX 25-FEB-1999.
XX
XX 07-AUG-1998; 98WO-GB02383.
XX
XX 13-AUG-1997; 97GB-0017192.
XX
XX (PLAN-) PLANT BIOSCIENCE LTD.
XX
XX Harberd NP, Peng J, Richards DE;

```

XX	
DR	WP1: 1999-181040/15.
XX	
PT	New Triticum Aestivum polynucleotides - encode a polypeptide which
PT	provides inhibition of the growth of plants, which inhibition is
PT	antagonised by gibberellin, used to confer a dwarf phenotype
XX	
PS	Disclosure: Fig 2b(5); 88pp; English.
XX	
CC	The specification describes polypeptides encoded by the Rht gene (and
CC	its homologues) that, when expressed in Triticum Aestivum, inhibit
CC	growth of the plant. This growth inhibition is antagonised by
CC	gibberellin. The products can be used to provide Rht expression in
CC	plants, conferring a dwarf phenotype on a plant which is correctable
CC	by treatment with gibberellin. In addition, the products can be
CC	used to produce Rht mutant plants which are dwarfed compared with
CC	wild-type, the dwarfing being gibberellin-insensitive. Taller plants
CC	may be made by knocking out Rht or the relevant homologous gene in
CC	the plant of interest. Plants may be made which are resistant to
CC	compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
CC	e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
CC	dwarf but let crop plants grow tall. The present sequence was
CC	obtained after partially sequencing wheat Rht clone 14a1.
XX	
SQ	Sequence 357 BP: 57 A; 125 C; 100 G; 60 T; 15 other;
	Query Match 5.0%; Score 107; DB 20; Length 357;
	Best Local Similarity 100.0%; Pred. No. 1,4e-36;
	Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1501 tactactccacatgttcgatctcctctcgaggcgagcagtccggcgggcccattccga 1560
Db	58 tactactccaccatgttgcgattccctcgaggcgagcagctccggcgggcccattccga 117
OY	1561 gtcctatcgggggcgtgcgtcgtctcctcgccgccgcgcgacgagacca 1607
Db	118 gtctatcgggggcgtgcgtcgtctcctcgccgccgcgcgacgagacca 164
	RESULT 14
	AAX36267
ID	AAX36267 standard; DNA: 230 BP.
XX	
AC	AAX36267;
DT	16-JUL-1999 (first entry)
XX	
DE	DNA sequence obtained after sequencing wheat Rht clone 14a1.
XX	
RW	Rht gene; homologue: Triticum aestivum; wheat; growth inhibition;
KW	antagonist: gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW	paclobutrazol; ss.
XX	
OS	Triticum aestivum.
XX	
PN	WO9909174-A1.
PD	25-FEB-1999.
XX	
PE	07-AUG-1998; 98WO-GB02383.
XX	
PR	13-AUG-1997; 97GB-0017192.
XX	
PA	(PLAN-) PLANT BIOSCIENCE LTD.
XX	
PI	Harberd NP, Peng J, Richards DE;
XX	
DR	WP1: 1999-181040/15.
XX	
PT	New Triticum Aestivum polynucleotides - encode a polypeptide which
PT	provides inhibition of the growth of plants, which inhibition is
PT	antagonised by gibberellin, used to confer a dwarf phenotype

[illegible]







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2002, 02:36:51 ; Search time 1804.19 Seconds  
(without alignments)  
12656.527 Million cell updates/sec

Title: US-09-485-529-14  
Perfect score: 2125  
Sequence: 1 atagagagagagagtagctc.....tgatgtgacgacgaactccg 2125

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlom:\*  
5: em\_estlpl:\*  
6: em\_estlba:\*  
7: em\_estlro:\*  
8: em\_estlov:\*  
9: em\_hlc:\*  
10: gb\_estl:\*  
11: gb\_est2:\*  
12: gb\_hlc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_trod:\*  
20: em\_gss\_vrl:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	4.1	400	10	BE414891 MML001.H0
2	77	3.6	268	11	BE201178 WHE0986_A
3	77	3.6	597	11	BE620181 HVSMEC001
4	68	3.2	778	11	BE308975 HVSMEC000
5	51	2.4	551	10	AT170444 606052F10
6	51	2.4	844	11	BE268018 HV_CEA001
7	46	2.2	559	11	BE618807 HVSMEC000
8	45	2.1	475	10	BE364489 P11_14_F0
9	45	2.1	481	11	BE417150 949053G08
10	45	2.1	527	10	BE595338 P11_48_G1
11	45	2.1	563	11	BE588097 FMI_37_F0
12	44	2.1	592	11	BE713881 EMI_60_F1

13	44	2.1	658	11	BE587541 FMI_37_F0
14	44	2.1	977	10	BE196286 HVSMEC009
15	44	2.1	981	10	BE454560 HVSMEC000
16	41	1.9	851	11	BE254247 HVSMEC000
17	38	1.8	440	11	BE109897 IP1_36_A1
18	35	1.6	228	10	BE095344 AU095344
19	35	1.6	263	11	D39460 RICS0803A R
20	35	1.6	388	11	C27475 C27475 Rice
21	35	1.6	399	10	AU091413 AU091413
22	35	1.6	416	11	BE051785 FMI_58_D0
23	35	1.6	443	11	BE655617 FMI_46_D1
24	35	1.6	467	10	AU222715 AU222715
25	35	1.6	596	10	AU065169 AU065169
26	29	1.4	259	11	BE411689 BE411689
27	29	1.4	427	10	A1577799 UT-R-ABO-
28	26	1.2	700	10	AL506960 AL506960
29	25	1.2	593	10	AL1881894 606074G08
30	24	1.1	373	10	A1967572 Lj1tripest
31	24	1.1	417	10	AV407563 AV407563
32	24	1.1	420	10	AV409142 AV409142
33	24	1.1	453	10	AM464360 BP230015B
34	24	1.1	566	10	AM066993 683012H10
35	24	1.1	586	11	BE1419686 LjNEST47e
36	23	1.1	223	10	AU009349 AU009349
37	23	1.1	223	10	AU009350 AU009350
38	23	1.1	283	10	AU007007 AU007007
39	23	1.1	283	10	AU007010 AU007010
40	23	1.1	353	10	AU006981 AU006981
41	23	1.1	358	10	AU008620 AU008620
42	23	1.1	358	10	AU008657 AU008657
43	23	1.1	360	10	AU013337 AU013337
44	23	1.1	361	11	D24385 R1CR1819A R
45	23	1.1	362	10	AU009468 AU009468

#### ALIGNMENTS

RESULT 1  
BE414891  
LOCUS MML001.H03P90430 ITRC MML Wheat Root Library Triticum aestivum  
DEFINITION CDNA clone MML001.H03, mRNA sequence.  
ACCESSION BE414891.1 GI:9412639  
VERSION BE414891  
KEYWORDS EST.  
SOURCE Triticum aestivum  
ORGANISM Triticum aestivum

REFERENCE  
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemelin,J.M., Jia,T., Joudrier,P., Langridge,P., Lazo,G.R., Lin,D.J., McGuire,P., Ogihara,T., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.  
International Triticaceae EST Cooperative (ITRC): Production of Expressed Sequence Tags for Species of the Triticaceae  
Unpublished (2000)  
Contact: Warburton M  
Applied Biotechnology Center, CIMMYT  
Apdo. Postal 6-641, 06600 Mexico DF MEXICO  
Tel: 52-5-7269091 ext 1381  
Fax: 52-5-7267558/59  
Email: mwarburton@cgmnet.com  
International Triticaceae EST Cooperative (ITRC)  
http://wheat.pw.usda.gov/genome.  
Location/Qualifiers

#### JOURNAL

COMMENT  
Applied Biotechnology Center, CIMMYT  
Apdo. Postal 6-641, 06600 Mexico DF MEXICO  
Tel: 52-5-7269091 ext 1381  
Fax: 52-5-7267558/59  
Email: mwarburton@cgmnet.com  
International Triticaceae EST Cooperative (ITRC)  
http://wheat.pw.usda.gov/genome.  
Location/Qualifiers

#### FEATURES

source  
1..400  
/organism="Triticum aestivum"  
/cultivar="Atlas"

```

/db.xref="taxon:4565"
/clone="MWL001.H03"
/clone_lib="TREC MWL Wheat Root Library"
/tissue_type="root"
/dev_stage="8 day old"
/note="Vector: pYES2 (Invitrogen); 0.5-1.5 kbp average insert size."

BASE COUNT      52 a      165 c      109 g      67 t      7 others
ORIGIN

Query Match          4.1%; Score 87; DB 10; Length 400;
Best Local Similarity 100.0%; Pred. No. 5.8e-27;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY    999 caaccagccatcttgtaggcgttcgcccgcgtgccgcgcgtgcaacgtcgtactcg 1058
      |||
Db     112 CAACGAGCCATCCTCGAGAGCGCTTCGCCGCCTGCGCGCGCATGCATGTGCAATTGG 171
      |||

OY    1059 catcaagcaggagatgcagtggcccg 1085
      |||
Db     172 CATCAAGCAGGGGATGATGAGTGCCCCG 198
      |||

RESULT 2
BF201178
LOCUS .. BF201178 268 bp mRNA EST 06-NOV-2000
DEFINITION WHE0986_A04.B0825 wheat pre-anthesis spike cDNA library Triticum aestivum CDNA clone WHE0986_A04.B08, mRNA sequence.
ACCESSION BF201178
VERSION .. BF201178.1 GI:1115968
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae ; Triticeae; Triticum.
1 (bases 1 to 268)
Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.U., Seaton,C.L. and Tong,J.C.
The structure and function of the expressed portion of the wheat genomes - Pre-anthesis spike cDNA library
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: StrataGene SK primer.

Location/Qualifiers
1..268
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE0986.A04.B08"
/clone_lib="wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; site_1: EcoRI; site_2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give plasmidic phagemids in the RT Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed In the OD

```

```

BASE COUNT      40 a      97 c      84 g      47 t
ORIGIN          Anderson lab (all other authors). "

Query Match            3.6%; Score 77; DB 11; Length 268;
Best Local Similarity 100.0%; Pred. No. 1,1e-22;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY    1345   gtcttcagatgacaccggcgtctcgccagaccgccgcaccttgagaaggctcttggcac 1404
       |||||
Db     18    GTCTTCGAGATGCACCGGGTGTCTCCGCCAGCCCCGCCCTTGAGAAAGTCTCTGGCAC 77
       |||||

OY    1405   gtgcgcgcgcgtgtggccc 1421
       |||||||
Db     78    GTGC GC GC GTG T GG CC 94

RESULT 3
BF620181 LOCUS BF620181 597 bp mRNA EST 22-FEB-2001
DEFINITION HVSMC0018M1lf Hordeum vulgare seedling shoot EST library
           HVCNDNA0003 (Etiolated and unstressed) Hordeum vulgare CDNA clone
           HVSMC0018M1lf, mRNA sequence.
BF620181 ACCESSION BF620181.2 GI:13109232
VERSION EST.
KEYWORDS Hordeum vulgare
SOURCE Barley.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
WOOD'R Development of a genetically and physically anchored EST resource
for barley genomics
Unpublished (2000)
JOURNAL On Dec 18, 2000 this sequence version replaced gi:11883915.
COMMENT Contact: Wing RA
           Clemson University Genomics Institute
           100 Jordan Hall, Clemson, SC 29654, USA
           Tel.: 864 656 7288
           Fax: 864 656 4293
           Email: rwing@clemson.edu
           Seq primer: AATTACCCTCATCAATAAGG
           High quality sequence stop: 564.
           Location/Qualifiers
               .size 597
               /organism="Hordeum vulgare"
               /cultivar="Morex"
               /db_xref="taxon:4513"
               /clone="HVSMC0018M1lf"
               /cdone_lip="hordeum vulgare seedling shoot EST library
HVCNDNA0003 (Etiolated and unstressed)"
               /tissue_type="Seedling shoot"
               /lab_host="TUJC121"
               /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/ To
order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT      99 a      202 c      204 g      91 t      1 others
ORIGIN
```

```

DB      288  GCGCGCGCGGACGAGCAGGTGATGTCGAGGTACTCGCGCGGAGATCTGCAAC 347
              |||
QY      1648  gtgtgtgctctgagagag 1664
              |||
DB      348  GTGTGTGCTGCGAGAGG 364
              |||

RESULT  4
BG308975  778 bp  mRNA  EST  22-FEB-2001
LOCUS     HG308975
DEFINITION HVSMC0001c18f Hordeum vulgare seedling shoot EST library
            HVCNA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone
            HVSMC0001c18f, mRNA sequence.
ACCESSION BG308975
VERSION   BG308975.1 GI:13109822
KEYWORDS  EST.
SOURCE    Hordeum vulgare
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
            ; Triticeae; Hordeum.
            1 (bases 1 to 778)
            Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
            Development of a genetically and physically anchored EST resource
            for barley genomics
            Unpublished (2000)
JOURNAL   Contact: Wing RA
COMMENT   Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAAACCTCAGCTAAAGG
            High quality sequence stop: 690.
            Location/Qualifiers
FEATURES  1..778
            source
            /organism="Hordeum vulgare"
            /cultivar="Morex"
            /db_xref="taxon:4513"
            /clone="HVSMEC0001C18f"
            /clone_lib="Hordeum vulgare seedling shoot EST library
            HVCNA0003 (Etiolated and unstressed)"
            /tissue_type="Seedling shoot"
            /lab_host="TJc121"
            /note="Vector: LambdaZAP, Site_1: EcoRI, Site_2: XhoI; For
            more details on library preparation and sequence analysis
            see http://www.genome.clemson.edu/projects/barley/ To
            order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT 96 a 295 c 261 g 124 t 2 others
ORIGIN

Query Match 3.2%; Score 68; DB 11; Length 778;
Best Local Similarity 100.0%; Pred. No. 9.3e-19;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DEFINITION 606052F10.x2 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION A1770444
VERSION   A1770444.1 GI:5268480
KEYWORDS  EST.
SOURCE    Zea mays.
ORGANISM  Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 551)
            Walbot,V.
            Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL   Unpublished (1999)
COMMENT   Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606052 row: F column: 10.
FEATURES  1..551
            source
            /organism="Zea mays"
            /cultivar="Ohio43"
            /db_xref="taxon:4577"
            /clone_lib="606 - Ear tissue cDNA library from Schmidt
            lab"
            /tissue_type="mixed"
            /dev_stage="ear length from 0.5 cm - 2.0 cm"
            /lab_host="XLOLR (Stratagene)"
            /note="Organ: Immature ear; Vector: pBK-CMV; Site_1: EcoRI
            ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
            lab"
BASE COUNT 66 a 213 c 184 g 88 t
ORIGIN

Query Match 2.4%; Score 51; DB 10; Length 551;
Best Local Similarity 100.0%; Pred. No. 1.9e-11;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: AATTAACCTCCTAAAGCG  
High quality sequence stop: 587.  
Location/Qualifiers

## FEATURES

source

1..844

/organism="Hordeum vulgare"  
/cultivar="C116155 (M1a13)"  
/db\_xref="taxon:4513"

/clone="HV\_CEA0019j17f"

/clone.lib="Hordeum vulgare seedling green leaf EST  
library HVCDNA0004 (Erysiphe infected & control)"  
/tissue.type="seedling green leaf"

/lab.host="TJUC121"  
/note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI; For  
more details on library preparation and sequence analysis  
see http://www.genome.clemson.edu/projects/barley/ To  
order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT 150 a 290 c 276 g 128 t  
ORIGIN

Query Match 2.4%; Score 51; DB 11; Length 844;  
Best Local Similarity 100.0%; Pred. No. 2e-11;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 178 tccgacatgcgagcgtgagcagaagctgagcagctcgagatggccatg 228  
|||||  
DB 299 TCCGACATGCGGACGTGGCGAAGCTGAGACGCTGAGATGGCCATG 349

## RESULT 7

BF618807

LOCUS BF618807 559 bp mRNA EST 22-FEB-2001  
DEFINITION HVMEC0007P09f Hordeum vulgare seedling shoot EST library  
HVCDNA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone  
HVMEC0007P09f, mRNA sequence.

ACCESSION BF618807  
VERSION BF618807.1 GI:1182541

KEYWORDS EST.  
SOURCE barley.  
ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
; Triticeae; Hordeum.

Wing,R., Close,T.J., Kleinholz,A., Wise,R., Begum,D., Fritsch,D., Yu  
,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo  
,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and  
Wood,T.

TITLE Development of a genetically and physically anchored EST resource  
for barley genomes  
JOURNAL Unpublished (2000)

COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu

Seq primer: AATTAACCTCCTAAAGCG  
High quality sequence start: 2  
High quality sequence stop: 405.  
Location/Qualifiers

## FEATURES

source

1..559

/organism="Hordeum vulgare"  
/cultivar="Morex"

/db\_xref="taxon:4513"  
/clone="HVMEC0007P09f"

/clone.lib="Hordeum vulgare seedling shoot EST library  
HVCDNA0003 (Etiolated and unstressed)"  
/tissue.type="seedling shoot"

/lab.host="TJUC121"

/note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI; For  
more details on library preparation and sequence analysis  
see http://www.genome.clemson.edu/projects/barley/ To  
order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT 100 a 172 c 197 g 90 t  
ORIGIN

Query Match 2.2%; Score 46; DB 11; Length 559;  
Best Local Similarity 100.0%; Pred. No. 2.7e-09;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1259 ccctgcgagcttgagccgttcacatgctcagccgagggcgagga 1304  
|||||  
DB 1 CGCTCGCGGACCTGGAGCCGTTTCATGCTGCACCGAGGCGAGGA 46

## RESULT 8

BE364489

LOCUS BE364489 475 bp mRNA EST 20-JUL-2000  
DEFINITION P11\_14\_F02.b1\_A002 Pathogen Induced 1 (P11) Sorghum bicolor cDNA,  
mRNA sequence.

ACCESSION BE364489  
VERSION BE364489.1 GI:9306046  
KEYWORDS EST.

SOURCE

Sorghum.

Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoidae; Andropogoneae; Sorghum.

1 (bases 1 to 475)  
Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt

,L.H.  
An EST database from Sorghum: pathogen-induced plants  
Unpublished (2000)

COMMENT Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmp@prattuga.edu

Sequences have been trimmed to exclude polyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.

Seq primer: JEN REV  
High quality sequence stop: 469  
POLYA-No.

Location/Qualifiers

## FEATURES

source

1..475

/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"

/clone.lib="Pathogen induced 1 (P11)"  
/note="Organ: Anthracnose-infected leaves from  
two-week-old sorghum plants 48 hr after inoculation;  
Vector: pBluescript II from Lambda Zap II; Site\_1: XhoI;  
Site\_2: EcoRI; Two-week-old sorghum plants (BTX 623  
cultivar) were infected with pathogen (isolate FRM421 of  
Colletotrichum graminicola, which is a sorghum isolate).  
RNA was prepared from infected leaves harvested from 45  
seedlings 48 hours after inoculation. Note: young  
seedlings (2 weeks old) exhibit juvenile resistant  
reaction, which is an incompatible interaction. As they  
grow older (4 weeks or older), plants resume susceptibility  
to anthracnose disease. The library was made from poly-A  
RNA in the cloning vector lambda Zap II. Clones to be  
sequenced were prepared by mass excision. WARNING: While  
most or all ESTs are expected to derive from the host  
plant, no effort was made to eliminate ESTs deriving from



DEFINITION	PM337_F08.bl_A003 Floral-Induced Meristem 1 (EM1) Sorghum
ACCESSION	BF588097
VERSION	BF588097.1
KEYWORDS	GI:11680421
SOURCE	EST.
ORGANISM	Sorghum propinquum.
REFERENCE	Sorghum propinquum
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
TITLE	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt,L.H.
JOURNAL	An EST database from Sorghum: floral-induced meristems
COMMENT	Unpublished (2000) Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. Seq primer: JEN REV High quality sequence stop: 477 POLYA-No.
FEATURES	location/Qualifiers
SOURCE	1..563 /organism="Sorghum propinquum" /db_xref="taxon:132711" /clone_id="Floral-Induced Meristem 1 (EM1)" /note="Organ: Floral-induced meristems; Vector: Bluescript II from lambda Zap II; Site.1: XhoI; Site.2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT	76 a 217 c 175 g 95 t
ORIGIN	
Query Match	2.18; Score 45; DB 11; Length 563;
Best Local Similarity	100.08; P-Id. No. 7.4e-09;
Matches	45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	730 gttgcacgcgctcgtcgctgcgagcgctgcagcaggaac 774
Dd	67 GTGCACGCCGCTCTGCGTGGCGGAGGCCGTGCACGACGAGAC 111
RESULT 12	
LOCUS	BG713881 592 bp mRNA EST 08-MAY-2001
DEFINITION	EM1_60_F10.g2_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
ACCESSION	BG713881
VERSION	BG713881.1
KEYWORDS	GI:14007831
SOURCE	EST.
ORGANISM	Sorghum bicolor
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
AUTHORS	Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
TITLE	An EST database from Sorghum: developing embryos
JOURNAL	Unpublished (2000)

COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmprratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. Seq primer: Polymix High quality sequence start: 9 High quality sequence stop: 563 POLYA-No.
FEATURES	Location/Qualifiers 1..592 /organism="Sorghum bicolor" /db_xref="taxon:4558" /clone_lib="Embryo 1 (EM1)" /note="Organ: Embryos germinated for 24 hr. Vector: plasmid II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the clonin vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT	99 a       179 c       184 g       130 t
ORIGIN	
Query Match	2.1%; Score 44; DB 11; Length 592;
Best Local Similarity	100.0%; Pred. No. 2e-08;
Matches 44; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY 1741	cacctggcctccaatgctaccagaacgcagcgacgctgtgtgc 1784       Db 107 CACCTGGGCTCCAATGCCTACAAGCAGCGACGACGTGTGTC 150
RESULT 13	
BF587541	658 bp mRNA EST 12-DEC-2000
LOCUS	F01.37.F08.g1.A003 Floral-induced Meristem 1 (F01) Sorghum
DEFINITION	Sorghum prolinquum cDNA, mRNA sequence.
ACCESSION	BF587541
VERSION	BF587541.1 GI:11679865
KEYWORDS	EST.
SOURCE	Sorghum prolinquum.
ORGANISM	Sorghum prolinquum. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 658) Cordonnier-Pratt,M.-W., Gingle,A., Sudman,M., Marsala,C. and Pratt, .L.H.
REFERENCE	An EST database from Sorghum: floral-induced meristems unpublished (2000)
AUTHORS	Contact: Cordonnier-Pratt MM
TITLE	Department of Botany
JOURNAL	The University of Georgia
COMMENT	Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmprratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. Seq primer: Polymix High quality sequence start: 35 High quality sequence stop: 658 POLYA-No.
FEATURES	Location/Qualifiers 1..658
SOURCE	/organism="Sorghum prolinquum" /db_xref="taxon:132711"



/clone.lib="floral-Induced Meristem 1 (FM1)"  
 /note="Organ: Floral-Induced Meristems; Vector:  
 pBluescript II from Lambda Zap II; Site.1: XhoI; Site.2:  
 EcoRI; mature plants were placed in a growth chamber for  
 15 days with 16 hr darkness and 8 hr light (flowering is  
 induced by short-day conditions); 16 days after being  
 returned to the greenhouse under natural long days during  
 late April/early May, meristems were harvested. The  
 library was made from poly-A RNA in the cloning vector  
 lambda Zap II. Clones to be sequenced were prepared by  
 mass excision."

BASE COUNT 115 a 196 c 205 g 142 t  
 ORIGIN

Query Match 2.1%; Score 44; DB 11; Length 658;  
 Best Local Similarity 100.0%; Pred. No. 2e-08;  
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1741 caccctggctccaatgcctacaagcagcagcagcgtctgagc 1784  
 ||||||||||||||||||||||||||||||||||||||||  
 Db 144 CACCTGGCTCCAAATGCTTACAGCAGCGACGACGCTGCTGCG 187

RESULT 14  
 BE196286 977 bp mRNA EST 02-MAR-2001  
 LOCUS HVSMH0091004f Hordeum vulgare 5-45 DAP spike EST library  
 DEFINITION HVCNDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMH0091004f,  
 mRNA sequence.

ACCESSION BE196286  
 VERSION BE196286  
 KEYWORDS GI:13188765

SOURCE  
 ORGANISM

Hordeum vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
 ; Triticeae; Hordeum.  
 1 (bases 1 to 977)  
 Wing, R., Close, T.J., Kleinof, A., Wise, R., Begum, D., Frisch, D., Yu  
 , Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo  
 , T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and  
 Wood, T.  
 Development of a genetically and physically anchored EST resource  
 for barley genomics  
 unpublished (2000)  
 On Jun 26, 2000 this sequence version replaced gi:8708481.

JOURNAL  
 COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: twing@clemson.edu  
 Seg primer: AATTACCTCACTAAAGG  
 High quality sequence stop: 618.

FEATURES  
 source

1..977  
 /organism="Hordeum vulgare"  
 /cultivar="Morex"  
 /db\_xref="taxon:4513"  
 /clone="HVSMH0091004f"  
 /clone.lib="Hordeum vulgare 5-45 DAP spike EST library  
 HVCNDA0009 (5 to 45 DAP)"  
 /tissue\_type="5-45 DAP Spike"  
 /lab\_host="SOLR"  
 /note="Vector: lambdaZAP; Site.1: EcoRI; Site.2: XhoI; For  
 more details on library preparation and sequence analysis  
 see http://www.genome.clemson.edu/projects/barley/ To  
 order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT 234 a 272 c 270 g 197 t 4 others  
 ORIGIN

Query Match 2.1%; Score 44; DB 10; Length 977;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-08;  
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1741 caccctggctccaatgcctacaagcagcagcagcgtctgagc 1784  
 ||||||||||||||||||||||||||||||||||||||||  
 Db 37 CACCTGGCTCCAAATGCTTACAGCAGCGACGACGCTGCTGCG 80

RESULT 15  
 BE454560 981 bp mRNA EST 02-MAR-2001  
 LOCUS HVSMH0094E04f Hordeum vulgare 5-45 DAP spike EST library  
 DEFINITION HVCNDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMH0094E04f,  
 mRNA sequence.

ACCESSION BE454560  
 VERSION BE454560  
 KEYWORDS GI:13189346

SOURCE  
 ORGANISM

Hordeum vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
 ; Triticeae; Hordeum.  
 1 (bases 1 to 981)  
 Wing, R., Close, T.J., Kleinof, A., Wise, R., Begum, D., Frisch, D., Yu  
 , Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo  
 , T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and  
 Wood, T.  
 Development of a genetically and physically anchored EST resource  
 for barley genomics  
 unpublished (2000)  
 On Jul 26, 2000 this sequence version replaced gi:9463094.

JOURNAL  
 COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: twing@clemson.edu  
 Seg primer: AATTACCTCACTAAAGG  
 High quality sequence stop: 672.  
 Location/Qualifiers  
 1..981

FEATURES  
 source

/organism="Hordeum vulgare"  
 /cultivar="Morex"  
 /db\_xref="taxon:4513"  
 /clone="HVSMH0094E04f"  
 /clone.lib="Hordeum vulgare 5-45 DAP spike EST library  
 HVCNDA0009 (5 to 45 DAP)"  
 /tissue\_type="5-45 DAP Spike"  
 /lab\_host="SOLR"  
 /note="Vector: lambdaZAP; Site.1: EcoRI; Site.2: XhoI; For  
 more details on library preparation and sequence analysis  
 see http://www.genome.clemson.edu/projects/barley/ To  
 order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT 277 a 244 c 241 g 190 t 29 others  
 ORIGIN

Query Match 2.1%; Score 44; DB 10; Length 981;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-08;  
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1741 caccctggctccaatgcctacaagcagcagcagcgtctgagc 1784  
 ||||||||||||||||||||||||||||||||||||||||  
 Db 18 CACCTGGCTCCAAATGCTTACAGCAGCGACGACGCTGCTGCG 61

Search completed: January 23, 2002., 04:03:30  
 Job time: 5199 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2002, 03:34:27 ; Search time 78.5 Seconds  
(without alignments)  
6130.774 Million cell updates/sec

Title: US-09-485-529-14

Perfect score: 2125  
Sequence: 1 atagagagagcaggtagctc.....tgatgtgacgacgactcgcg 2125

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 11323899 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents.NA:\*  
1: /cgn2\_6/prodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/PCRTOS.COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/ackfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	1.0	4308	3	US-08-938-291A-8
2	21	1.0	1491	4	US-09-082-092-9
3	20	0.9	590	1	US-08-314-309A-10
4	20	0.9	1028	4	US-08-118-200-1
5	20	0.9	1028	4	US-08-458-745-1
6	20	0.9	1461	1	US-08-403-634-1
7	20	0.9	1461	4	US-08-913-441B-1
8	20	0.9	2036	3	US-08-923-454A-17
9	20	0.9	2205	3	US-08-888-077A-41
10	20	0.9	3172	1	US-08-314-309A-1
11	20	0.9	3517	2	US-08-642-406A-20
12	20	0.9	3517	4	US-08-434-000A-1
13	20	0.9	3517	4	US-09-312-157-1
14	20	0.9	9997	1	US-08-246-982A-15
15	20	0.9	9997	1	US-08-453-265-15
16	20	0.9	10103	2	US-08-457-273B-7
17	20	0.9	10348	2	US-08-457-273B-41
18	20	0.9	10348	3	US-08-556-419-13
19	20	0.9	10348	4	US-09-041-886-14
20	20	0.9	10366	1	US-08-246-982A-5
21	20	0.9	10366	1	US-08-453-265-5
22	19	0.9	71	2	US-08-712-948-4
23	19	0.9	472	2	US-08-465-380-33
24	19	0.9	472	2	US-08-486-397-33
25	19	0.9	472	2	US-08-486-399-33
26	19	0.9	472	2	US-08-461-965-33
27	19	0.9	472	2	US-08-634-641-33

c 28	19	0.9	472	3	US-09-249-471-33	Sequence 33, Appl
c 29	19	0.9	472	3	US-09-249-472-33	Sequence 33, Appl
c 30	19	0.9	472	3	US-09-249-451-33	Sequence 33, Appl
c 31	19	0.9	472	3	US-08-809-455-33	Sequence 33, Appl
c 32	19	0.9	472	3	US-09-249-461-33	Sequence 33, Appl
c 33	19	0.9	472	3	US-09-249-448-33	Sequence 33, Appl
c 34	19	0.9	472	2	US-08-465-380-35	Sequence 35, Appl
c 35	19	0.9	472	2	US-08-486-397-35	Sequence 35, Appl
c 36	19	0.9	477	2	US-08-486-399-35	Sequence 35, Appl
c 37	19	0.9	477	2	US-08-461-965-35	Sequence 35, Appl
c 38	19	0.9	477	2	US-08-634-641-35	Sequence 35, Appl
c 39	19	0.9	477	3	US-09-249-471-35	Sequence 35, Appl
c 40	19	0.9	477	3	US-09-249-472-35	Sequence 35, Appl
c 41	19	0.9	477	3	US-09-249-451-35	Sequence 35, Appl
c 42	19	0.9	477	3	US-08-809-455-35	Sequence 35, Appl
c 43	19	0.9	477	3	US-09-249-461-35	Sequence 35, Appl
c 44	19	0.9	477	3	US-09-249-448-35	Sequence 35, Appl
c 45	19	0.9	3358	2	US-08-469-537A-104	Sequence 104, App

#### ALIGNMENTS

RESULT 1  
US-08-938-291A-8/c  
; Sequence 8, Application US/08938291A  
; Patent No. 6117673  
; GENERAL INFORMATION:  
; APPLICANT: Lev, Sima  
; APPLICANT: Plovman, Gregory D.  
; APPLICANT: Schlessinger, Joseph  
; TITLE OF INVENTION: RDB PROTEINS AND RELATED  
; TITLE OF INVENTION: PRODUCTS AND METHODS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FastSeq  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/938,291A  
; FILING DATE: September 26, 1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/027,337  
; FILING DATE: October 11, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 228/172  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELETYPE: 67-3510  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4308 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-938-291A-8  
Query Match 1.0%; Score 22; DB 3; Length 4308;

Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 581 cctgcgtcgtcctcctcctc 602  
|||||  
Db 1420 CCTGCTGTCATCTCTCTC 1399

RESULT 2  
US-09-082-092-9  
; Sequence 9, Application US/09082092  
; Patent No. 6251628

GENERAL INFORMATION:

APPLICANT: NAKAO, Atsuhito  
APPLICANT: MOREN, Anita  
APPLICANT: Heuchel, Rainer  
APPLICANT: Itoh, Susumu  
APPLICANT: Aitakhte, Mozghan  
APPLICANT: Souchevnytskyi, Serhiy  
APPLICANT: Landstrom, Marene  
APPLICANT: Heldin, Nils-Erik  
APPLICANT: Heldin, Carl-Henrik  
APPLICANT: ten Dijke, Peter

TITLE OF INVENTION: SMAD7 AND USES THEREOF  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA

COUNTRY: U.S.A.  
ZIP: 02210-2211

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/082,092  
FILING DATE: 20-MAY-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/047,221  
FILING DATE: 20-MAY-1997

APPLICATION NUMBER: 60/060,465  
FILING DATE: 30-SEP-1997

APPLICATION NUMBER: 60/075,940  
FILING DATE: 25-FEB-1998

APPLICATION NUMBER: 60/077,033  
FILING DATE: 06-MAR-1998

ATTORNEY/AGENT INFORMATION:  
NAME: Van Amsterdam, John R.

REGISTRATION NUMBER: 40,212  
REFERENCE/DOCKET NUMBER: L0461/7032

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441

TELEX:

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 1491 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA  
US-09-082-092-9

Query Match 1.0%; Score 21; DB 4; Length 1491;  
Best Local Similarity 100.0%; Pred. No. 4.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 58 agcggcggcgcgcgtgacgc 78  
|||||  
Db 79 AGCGGCGGCGCGGTGCGCGC 99

RESULT 3  
US-08-314-309A-10/c  
; Sequence 10, Application US/08314309A  
; Patent No. 5677141

GENERAL INFORMATION:

APPLICANT: ISOGAI, TAKAO  
APPLICANT: FUKAGAWA, MASAO  
APPLICANT: IWAMI, MORITA  
APPLICANT: ARAMORI, ICHIRO  
APPLICANT: KOJO, HITOSHI  
TITLE OF INVENTION: PROCESS FOR PRODUCING 7-AMINOCEPHEM  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:

ADDRESSEE: P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/314,309A  
FILING DATE: 30-SEP-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/631,906  
FILING DATE: 21-DEC-1990

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 5677141man F.

REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-863-0 CONT

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 590 base pairs

TYPE: nucleic acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)  
US-08-314-309A-10

Query Match 0.9%; Score 20; DB 1; Length 590;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 123 ggaagggaaggagtgacgc 142  
|||||  
Db 153 GGAGGGGAGGAGGTGAGCG 134

RESULT 4

US-08-118-200-1  
; Sequence 1, Application US/08118200  
; Patent No. 6197500

GENERAL INFORMATION:

APPLICANT: SUTHERLAND, Grant R  
APPLICANT: RICHARDS, Robert I

APPLICANT: SCHLESSINGER, David

APPLICANT: NAGARAJA, Ramalah  
APPLICANT: KREMER, Eric J  
APPLICANT: YU, Sui  
APPLICANT: BAKER, Elizabeth  
APPLICANT: MULLEY, John C  
APPLICANT: MANDEL, Jean-Louis  
APPLICANT: PRITCHARD, Melanie April  
APPLICANT: LYNCH, Michael  
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED  
NUMBER OF SEQUENCES: 13  
TITLE OF INVENTION: FRAGILE X SYNDROME  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/118,200  
FILING DATE: 09-SEP-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/802,650  
FILING DATE: 05-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,232  
FILING DATE: 20-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/638,518  
FILING DATE: 04-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/966,517  
FILING DATE: 23-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 020160-164  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1028 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-118-200-1

Query Match 0.9%; Score 20; DB 4; Length 1028;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 gcggcgagcgagtgagcggc 78  
|||||

Db 396 GCGGCGGCGCGGTGCGGC 415

RESULT 5  
US-08-458-745-1  
; Sequence 1, Application US/08458745  
; Patent No. 6242576  
; GENERAL INFORMATION:  
; APPLICANT: SUTHERLAND, Grant R  
; APPLICANT: RICHARDS, Robert I  
; APPLICANT: SCHLESSINGER, David

APPLICANT: NAGARAJA, Ramalah  
APPLICANT: KREMER, Eric J  
APPLICANT: YU, Sui  
APPLICANT: BAKER, Elizabeth  
APPLICANT: MULLEY, John C  
APPLICANT: MANDEL, Jean-Louis  
APPLICANT: PRITCHARD, Melanie April  
APPLICANT: LYNCH, Michael  
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED  
NUMBER OF SEQUENCES: 13  
TITLE OF INVENTION: FRAGILE X SYNDROME  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,745  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/118,200  
FILING DATE: 09-SEP-1993  
APPLICATION NUMBER: US 07/802,650  
FILING DATE: 05-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,232  
FILING DATE: 20-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/638,518  
FILING DATE: 04-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/966,517  
FILING DATE: 23-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 020160-164  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1028 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-458-745-1

Query Match 0.9%; Score 20; DB 4; Length 1028;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 gcggcgagcgagtgagcggc 78  
|||||

Db 396 GCGGCGGCGCGGTGCGGC 415

RESULT 6  
US-08-403-634-1/c  
; Sequence 1, Application US/08403634  
; Patent No. 5674748  
; GENERAL INFORMATION:  
; APPLICANT: Giordano, Antonio

```

; TITLE OF INVENTION: NOVEL HUMAN CYCLIN-DEPENDENT
; TITLE OF INVENTION: KINASE-LIKE PROTEINS AND METHODS
; TITLE OF INVENTION: OF USING THE SAME
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESS: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESS: No. 56747815
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,634
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/208,575
; FILING DATE: 08-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1482
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1461 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 65..1180
; US-08-403-634-1

Query Match      0.9%; Score 20; DB 1; Length 1461;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 caacgcgcgcgcgcgcgcgc 373
DB 55 CAACGCGCGCGCGCGCGCCC 36

RESULT 7
US-08-913-441B-1/c
; Sequence 1, Application US/08913441B
; Patent No. 6162612
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: No. 6162612el Human Cyclin-Dependent Kinase-Like Proteins and
; TITLE OF INVENTION: Methods of Using The Same
; FILE REFERENCE: 8321-76 C11
; CURRENT APPLICATION NUMBER: US/08/913,441B
; CURRENT FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 08/403,634
; PRIOR FILING DATE: 1995-03-14
; PRIOR APPLICATION NUMBER: PCT/US96/03557
; PRIOR FILING DATE: 1996-03-14
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1461
; TYPE: DNA
```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (65)..(1180)
; US-08-913-441B-1

Query Match      0.9%; Score 20; DB 4; Length 1461;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 caacgcgcgcgcgcgcgcgc 373
DB 55 CAACGCGCGCGCGCGCGCCC 36

RESULT 8
US-08-923-454A-17/c
; Sequence 17, Application US/08923454A
; Patent No. 6004794
; GENERAL INFORMATION:
; APPLICANT: Creasy, Caretha
; APPLICANT: Livt, George
; APPLICANT: Kairan, Eric
; APPLICANT: Clinkenbeard, Helen
; APPLICANT: Browne, Michael
; APPLICANT: Southan, Christopher
; TITLE OF INVENTION: HUMAN SERINE PROTEASE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESS: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,454A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/025436
; FILING DATE: 06-SEPT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50547
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2036 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; MOLECULE TYPE: CDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORIGINAL SOURCE: Feature polymorphism at 1325
; US-08-923-454A-17

Query Match      0.9%; Score 20; DB 3; Length 2036;
Best Local Similarity 100.0%; Pred. No. 11;
```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 ggcggcgagcggtgagcgc 78  
 |||  
 Db 38 ggcggcgagcggtgagcgc 19

## RESULT 9

US-08-888-077A-41/C  
 ; Sequence 41, Application US/08888077A  
 ; Patent No. 6020143  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ST. GEORGE-HYSLOP, PETER H  
 ; APPLICANT: ROMMENS, JOHANNA M  
 ; APPLICANT: FRASER, PAUL E  
 ; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
 ; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE AND USBS THEREFOR.  
 ; NUMBER OF SEQUENCES: 41  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: LERNER, DAVID, LITTENBERG, KROMHOLZ & MENTLIK  
 ; STREET: 600 SOUTH AVENUE WEST  
 ; CITY: WESTFIELD  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 07090-1497  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/888,077A  
 ; FILING DATE: 03-JUL-1997  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/592,541  
 ; FILING DATE: 26-JAN-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: PALISI, THOMAS M  
 ; REGISTRATION NUMBER: 36,629  
 ; REFERENCE/DOCKET NUMBER: SCHERING 3.0-017 CIP CIP CIP IV  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (908) 654-5000  
 ; TELEFAX: (908) 654-7866  
 ; INFORMATION FOR SEQ ID NO: 41:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2205 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: misc-feature  
 ; LOCATION: 1..2205  
 ; OTHER INFORMATION: /note="mutm1-TM2"

Query Match 0.9%; Score 20; DB 3; Length 2205;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 ggcggcgagcggtgagcgc 78  
 |||  
 Db 113 ggcggcgagcggtgagcgc 94

## RESULT 10

US-08-314-309A-1/C  
 ; Sequence 1, Application US/08314309A  
 ; Patent No. 5677141  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ISOGAI, TAKAO  
 ; APPLICANT: FUKAGAWA, MASAO

APPLICANT: IWAMI, MORITA  
 APPLICANT: ARAMORI, ICHIRO  
 APPLICANT: KOJO, HITOSHI  
 TITLE OF INVENTION: PROCESS FOR PRODUCING 7-AMINOCEPHEM  
 TITLE OF INVENTION: COMPOUND OR SALTS THEREOF  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
 ADDRESS: P.C.  
 STREET: 1755 S. Jefferson Davis Highway, Suite 400  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/314,309A  
 FILING DATE: 30-SEP-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/631,906  
 FILING DATE: 21-DEC-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: OBLON, No. 5677141man F.  
 REGISTRATION NUMBER: 24,618  
 REFERENCE/DOCKET NUMBER: 18-863-0 CONT  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 413-3000  
 TELEFAX: (703) 413-2220  
 TELEX: 248655 OPAT UR  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3172 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: unknown  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA (genomic)  
 US-08-314-309A-1

Query Match 0.9%; Score 20; DB 1; Length 3172;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 ggcggcgagcggtgagcgc 142  
 |||  
 Db 484 ggcggcgagcggtgagcgc 465

## RESULT 11

US-08-642-406A-20/C  
 ; Sequence 20, Application US/08642406A  
 ; Patent No. 5959177  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hein, Mich B.  
 ; APPLICANT: Hatt, Andrew C.  
 ; APPLICANT: Ma, Julian K.C.  
 ; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED  
 ; TITLE OF INVENTION: SECRETORY ANTIBODIES  
 ; NUMBER OF SEQUENCES: 26  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE  
 ; STREET: 10666 NO. 5959177th Torrey Pines Road, TPC-8  
 ; CITY: La Jolla  
 ; STATE: California  
 ; COUNTRY: US  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk





```

:      SEQUENCE CHARACTERISTICS:
:      LENGTH: 3517 base pairs
:      TYPE: nucleic acid
:      STRANDEDNESS: single
:      TOPOLOGY: linear
:
:      FEATURE:
:      NAME/KEY: Coding Sequence
:      LOCATION: 124....2445
:      SEQUENCE DESCRIPTION: SEQ ID NO: 1:
:
US-09-312-157-1

```

	Query Match	0.9%;	Score 20;	DB 4;	Length 3517;
	Best Local Similarity	100.0%	Pred. No. 11;		
Matches	20; Conservative	0;	Mismatches	0;	Indels
					Gaps
QY	59 ggcggcgccgctgtgcgcgc	78			
DB	2486 gcgcgcgcgcgcgcgcgcgc	2467			

RESULT 14  
 US-08-246-982A-15/c  
 Sequence 15 Application US/08246982A  
 Patent No. 5686288  
 GENERAL INFORMATION:  
 APPLICANT: Macdonald, Marcy E.  
 APPLICANT: Ambrose, Christine M.  
 APPLICANT: Duyao, Mabel P.  
 APPLICANT: Gussella, James F.  
 TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof  
 NUMBER OF SEQUENCES: 25  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
 STREET: 1100 New York Avenue  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: U.S.A.  
 ZIP: 20005  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08246,982A  
 FILING DATE: May 20, 1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Goldstein, Jorge, A.  
 REGISTRATION NUMBER: 29,021  
 REFERENCE/DOCKET NUMBER: 0609.3880002  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 371-2600  
 TELEFAX: (202) 371-2540  
 INFORMATION FOR SEQ ID NO: 15:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9997 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: both  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 90..9446  
 US-08-246-982A-15

Query Match	0.98;	Score 20;	DB 1;	Length 9997;
Best Local Similarity	100.0%;	Pred. No. 9.5;		
Matches	20;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

Db 195 GCGGCGGCGGCGGTGGCGC 176

RESULT 15  
US-08-453-265-15/c  
Sequence 15, Application US/08453265  
Patent No. 5693757  
GENERAL INFORMATION:  
APPLICANT: Macdonald, Marcy E.  
APPLICANT: Ambrose, Christine M.  
APPLICANT: Duyao, Mabel P.  
APPLICANT: Gusella, James F.  
TITLE OF INVENTION: Huntington DNA, Protein And Uses Thereof  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,265  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Ludwig, Steven R.  
REGISTRATION NUMBER: 36,203  
REFERENCE/DOCKET NUMBER: 0609,3880003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9997 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 90..9446  
US-08-453-265-15

	Query Match	0.9%; Best Local Similarity	Score 20; 100.0%	DB 1; Pred. No. 9.5;	Length 997;
Matches	20;	Conservative	0;	Mismatches	0;
Indels					0;
Gaps					0;
QY	59	gcggcgccgagcgtgacgagc	78		
Db	195	gcggcgccgagcgtgacgagc	176		

Search completed: January 23, 2002, 05:00:59  
Job time: 5192 sec

---



XX PS Disclosure; Fig 8b; 88pp; English.  
CC CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in *Triticum aestivum*, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence is encoded by  
CC the wheat Rht clone 5a1 genomic sequence.  
XX SQ Sequence 623 AA;

Query Match 100.0%; Score 13; DB 20; Length 623;  
Best Local Similarity 100.0%; Pred. No. 0.00044;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNAPPPPLPPAPQ 13  
DB 109 LNAPPPPLPPAPQ 121  
|||||

RESULT 2  
AAG27119  
ID AAG27119 standard; Protein; 33 AA.  
XX AC AAG27119;  
XX DT 17-OCT-2000 (first entry)  
XX DE Zea mays protein fragment SEQ ID NO: 31833.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence; corn.  
XX OS Zea mays subsp. mays.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131448.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 01-JUN-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137502.  
PR 08-JUN-1999; 99US-0137724.  
PR 10-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139730.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 26-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145951.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 69.2%; Score 9; DB 21; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
 Db 13 PPPPLPPAP 21

RESULT 3  
 AAB80893 standard; protein; 182 AA.  
 XX  
 AC AAB80893;  
 XX  
 DT 30-MAY-2001 (first entry)  
 XX  
 DE Human ATFX leucine zipper domain.  
 XX  
 KW Human; ATFX; gamma aminobutyric acid B receptor; GABA B receptor;  
 KM transcription factor; leucine zipper.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200116596-A2.  
 XX  
 PD 08-MAR-2001.  
 XX  
 PF 31-AUG-2000; 2000MO-GB03361.  
 XX  
 PR 31-AUG-1999; 99GB-0020569.  
 PR 12-JAN-2000; 2000GB-0000516.  
 XX  
 PA (GLAX ) GLAXO GROUP LTD.  
 XX  
 PI White J, Wise A, Marshall F;  
 XX  
 DR WPI; 2001-265904/27.  
 DR N-PSDB; AAF77878.  
 XX  
 PT Identifying modulators of gamma aminobutyric acid(GABA)-B receptor  
 PT mediated activity by monitoring the interaction between GABAB receptor  
 PT and the CREB/ATF transcription factors in the presence of a test  
 PT substance  
 XX  
 PS Claim 14; Page 64-65; 71pp; English.  
 XX  
 CC The present invention relates to a method for identifying a modulator of  
 CC gamma aminobutyric acid,B (GABA\_B) receptor-mediated activity, by  
 CC monitoring the interaction between a CREB/ATF transcription factor  
 CC capable of binding to GABA\_B receptor. The present sequence is the  
 CC leucine zipper domain of human ATFX (see AAB80892), which was used in the  
 CC method of the present invention. ATFX is a member of the CREB/ATF family  
 CC of transcription factors. ATFX contains a bZIP domain, by which it  
 CC interacts with the coiled coil domain of GABA\_B receptor. Modulators of  
 CC GABA\_B receptor activity are useful for treating central nervous system  
 CC or peripheral nervous system disorders.  
 XX  
 SQ Sequence 182 AA;

Query Match 69.2%; Score 9; DB 22; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 0.45;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
 |||||  
 DB 31 PPPPLPPAP 39

RESULT 4  
 AAB80892  
 ID AAB80892 standard; protein; 216 AA.

AC AAB80892;  
 XX  
 DT 30-MAY-2001 (first entry)

DE Human ATFX.

KW Human: ATFX; gamma aminobutyric acid B receptor; GABA B receptor;  
 KW transcription factor.

OS Homo sapiens.

PN WO200116596-A2.

PD 08-MAR-2001.

PF 31-AUG-2000; 2000WO-GB03361.

PR 31-AUG-1999; 99GB-0020569.

PR 12-JAN-2000; 2000GB-0000516.

PA (GLAXO ) GLAXO GROUP LTD.

PI White J, Wise A, Marshall F;

DR MPI: 2001-265904/27.

DR N-PSDB; AAF77877.

XX Identifying modulators of gamma aminobutyric acid(GABA)-B receptor  
 PT mediated activity by monitoring the interaction between GABAB receptor  
 PT and the CREB/ATF transcription factors in the presence of a test  
 PT substance

PS Claim 14; Fig 1; 71pp; English.

XX The present invention relates to a method for identifying a modulator of  
 CC gamma aminobutyric acid,B (GABA,B) receptor-mediated activity, by  
 CC monitoring the interaction between a CREB/ATF transcription factor  
 CC capable of binding to GABA,B receptor. The present sequence is human  
 CC ATFX, which was used in the method of the present invention. ATFX is a  
 CC member of the CREB/ATF family of transcription factors. ATFX contains a  
 CC bZIP domain, by which it interacts with the coiled coil domain of GABA,B  
 CC receptor. Modulators of GABA,B receptor activity are useful for treating  
 CC central nervous system or peripheral nervous system disorders.

XX Sequence 216 AA;

Query Match 69.2%; Score 9; DB 22; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 0.51;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
 |||||  
 DB 65 PPPPLPPAP 73

RESULT 5  
 AAW21733

ID AAW21733 standard; Protein; 223 AA.

XX AAW21733;

AC AAW21733;  
 XX  
 DT 01-OCT-1997 (first entry)

DE NIP-1 encoded by clone 59.

KW NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;  
 KW cell division; proliferation; antibody; Ab; detection;  
 KW malignant cell growth.

OS Homo sapiens.

PN WO9640917-A1.

PD 19-DEC-1996.

PF 07-JUN-1996; 96WO-US09504.

PR 07-JUN-1995; 95US-0478408.

PA (UYVIA ) UNIV YALE.

PI McPherson SMG, Snyder MP;

DR MPI: 1997-077270/07.

DR N-PSDB; AAT77784.

XX New nucleic acid encoding nuclear mitotic appts. interacting  
 PT proteins - useful for modulating cell division and proliferation and  
 PT in diagnosis

PS Claim 28; Page 57-58; 78pp; English.

XX The sequences given in AAW21733-35 represent NuMA binding proteins (NuMA  
 CC = nuclear mitotic apparatus). These protein sequences were identified  
 CC using the fusion proteins given in AAW21731-32. Compounds which  
 CC interfere with the interaction of NuMA with a known NIP (NuMA interacting  
 CC protein) are used to modulate cell division and/or proliferation. Ab,  
 CC raised conventionally using NIP-1 or -2 as immunogen, are used to detect  
 CC NIP (or their complexes) and to block their activity for diagnostic or  
 CC therapeutic use, e.g. to detect defective NuMA or NIP which may be  
 CC markers for aberrant (including malignant) cell growth (which can also  
 CC be detected by nucleic acid sequencing). Also where malignancy is  
 CC related to defects in NuMA or NIP, it can be treated by administration  
 CC of the appropriate functional protein. This protein is rich in  
 CC proline residues.

XX Sequence 223 AA;

Query Match 69.2%; Score 9; DB 18; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 0.53;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
 |||||  
 DB 2 PPPPLPPAP 10

RESULT 6

AAB95859  
 ID AAB95859 standard; Protein; 282 AA.

XX AAB95859;

AC AAB95859;  
 XX  
 DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:18922.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX

OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 18922; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 SQ Sequence 282 AA;  
 XX  
 QY  
 Db 4 PPPPLPPAP 12  
 131 PPPPLPPAP 139  
 XX  
 RESULT 7  
 AAB80896  
 ID AAB80896 standard; protein; 282 AA.  
 XX  
 AC AAB80896;  
 XX  
 DT 30-MAY-2001 (first entry)  
 XX  
 DE Human ATRX with N-terminal extension.  
 XX

KM Human; ATRX; gamma aminobutyric acid B receptor; GABA B receptor;  
 KM transcription factor.  
 XX  
 XX Homo sapiens.  
 OS  
 PN WO200116596-A2.  
 XX  
 PD 08-MAR-2001.  
 XX  
 PF 31-AUG-2000; 2000WO-GB03361.  
 XX  
 PR 31-AUG-1999; 99GB-0020569.  
 PR 12-JAN-2000; 2000GB-0000516.  
 XX  
 PA (GLAXO) GLAXO GROUP LTD.  
 XX  
 PI White J, Wise A, Marshall F;  
 PI N-PSDB; AAF77880.  
 XX  
 DR WPI; 2001-265904/27.  
 XX  
 PT Identifying modulators of gamma aminobutyric acid(GABA)-B receptor  
 PT mediated activity by monitoring the interaction between GABAB receptor  
 PT and the CREB/ATF transcription factors in the presence of a test  
 PT substance -  
 XX  
 PS Claim 14; Fig 9; 71pp; English.  
 XX  
 CC The present invention relates to a method for identifying a modulator of  
 CC gamma aminobutyric acid<sub>B</sub> (GABA<sub>B</sub>) receptor-mediated activity, by  
 CC monitoring the interaction between a CREB/ATF transcription factor  
 CC capable of binding to GABA<sub>B</sub> receptor. The present sequence is human  
 CC ATRX, which was used in the method of the present invention. ATRX is a  
 CC member of the CREB/ATF family of transcription factors. ATRX contains a  
 CC bZIP domain, by which it interacts with the coiled coil domain of GABA<sub>B</sub>  
 CC receptor. Modulators of GABA<sub>B</sub> receptor activity are useful for treating  
 CC central nervous system or peripheral nervous system disorders. The  
 CC present sequence has an N-terminal extension compared to AAB80892.  
 CC  
 SQ Sequence 282 AA;  
 XX  
 QY  
 Db 4 PPPPLPPAP 12  
 131 PPPPLPPAP 139  
 XX  
 RESULT 8  
 AAG07315  
 ID AAG07315 standard; Protein; 183 AA.  
 XX  
 AC AAG07315;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 4422.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX

PR	25-FEB-1999;	99US-0121825;
PR	05-MAR-1999;	99US-0123180;
PR	09-MAR-1999;	99US-0123548;
PR	23-MAR-1999;	99US-0125788;
PR	28-MAR-1999;	99US-0126264;
PR	29-MAR-1999;	99US-0126785;
PR	01-APR-1999;	99US-0127462;
PR	06-APR-1999;	99US-0128234;
PR	08-APR-1999;	99US-0128714;
PR	16-APR-1999;	99US-0129845;
PR	19-APR-1999;	99US-0130047;
PR	21-APR-1999;	99US-0130477;
PR	23-APR-1999;	99US-0130510;
PR	23-APR-1999;	99US-0130891;
PR	28-APR-1999;	99US-0131449;
PR	30-APR-1999;	99US-0132048;
PR	30-APR-1999;	99US-0132407;
PR	04-MAY-1999;	99US-0132484;
PR	05-MAY-1999;	99US-0132485;
PR	06-MAY-1999;	99US-0132486;
PR	07-MAY-1999;	99US-0132487;
PR	07-MAY-1999;	99US-0132865;
PR	11-MAY-1999;	99US-0134256;
PR	14-MAY-1999;	99US-0134218;
PR	14-MAY-1999;	99US-0134219;
PR	14-MAY-1999;	99US-0134221;
PR	14-MAY-1999;	99US-0134370;
PR	18-MAY-1999;	99US-0134768;
PR	19-MAY-1999;	99US-0134941;
PR	20-MAY-1999;	99US-0135121;
PR	21-MAY-1999;	99US-0135353;
PR	24-MAY-1999;	99US-0135629;
PR	25-MAY-1999;	99US-0136021;
PR	27-MAY-1999;	99US-0136392;
PR	28-MAY-1999;	99US-0136782;
PR	01-JUN-1999;	99US-0137222;
PR	03-JUN-1999;	99US-0137528;
PR	04-JUN-1999;	99US-0137502;
PR	07-JUN-1999;	99US-0137724;
PR	08-JUN-1999;	99US-0138094;
PR	10-JUN-1999;	99US-0138540;
PR	10-JUN-1999;	99US-0138847;
PR	14-JUN-1999;	99US-0139119;
PR	16-JUN-1999;	99US-0139452;
PR	16-JUN-1999;	99US-0139453;
PR	17-JUN-1999;	99US-0139492;
PR	18-JUN-1999;	99US-0139454;
PR	18-JUN-1999;	99US-0139455;
PR	18-JUN-1999;	99US-0139456;
PR	18-JUN-1999;	99US-0139457;
PR	18-JUN-1999;	99US-0139458;
PR	18-JUN-1999;	99US-0139459;
PR	18-JUN-1999;	99US-0139460;
PR	18-JUN-1999;	99US-0139461;
PR	18-JUN-1999;	99US-0139462;
PR	18-JUN-1999;	99US-0139463;
PR	18-JUN-1999;	99US-0139750;
PR	18-JUN-1999;	99US-0139763;
PR	21-JUN-1999;	99US-0139817;
PR	22-JUN-1999;	99US-0139889;
PR	23-JUN-1999;	99US-0139899;
PR	23-JUN-1999;	99US-0140353;
PR	24-JUN-1999;	99US-0140354;
PR	24-JUN-1999;	99US-0140695;
PR	28-JUN-1999;	99US-0140820;
PR	29-JUN-1999;	99US-0140991;
PR	30-JUN-1999;	99US-0141287;
PR	01-JUL-1999;	99US-0141842;
PR	01-JUL-1999;	99US-0142154;
PR	02-JUL-1999;	99US-0142055;
PR	06-JUL-1999;	99US-0142390;
PR	08-JUL-1999;	99US-0142803;
PR	09-JUL-1999;	99US-0142920;
PR	12-JUL-1999;	99US-0142927;
PR	13-JUL-1999;	99US-0143542;
PR	14-JUL-1999;	99US-0143624;
PR	15-JUL-1999;	99US-0144005;
PR	16-JUL-1999;	99US-0144085;
PR	16-JUL-1999;	99US-0144086;
PR	19-JUL-1999;	99US-0144331;
PR	19-JUL-1999;	99US-0144332;
PR	19-JUL-1999;	99US-0144332;
PR	19-JUL-1999;	99US-0144334;
PR	19-JUL-1999;	99US-0144334;
PR	20-JUL-1999;	99US-0144335;
PR	20-JUL-1999;	99US-0144632;
PR	20-JUL-1999;	99US-0144884;
PR	21-JUL-1999;	99US-0144814;
PR	21-JUL-1999;	99US-0145086;
PR	21-JUL-1999;	99US-0145087;
PR	22-JUL-1999;	99US-0145087;
PR	22-JUL-1999;	99US-0145089;
PR	22-JUL-1999;	99US-0145192;
PR	23-JUL-1999;	99US-0145145;
PR	23-JUL-1999;	99US-0145218;
PR	23-JUL-1999;	99US-0145224;
PR	26-JUL-1999;	99US-0145276;
PR	27-JUL-1999;	99US-0145913;
PR	27-JUL-1999;	99US-0145918;
PR	27-JUL-1999;	99US-0145919;
PR	28-JUL-1999;	99US-0145951;
PR	02-AUG-1999;	99US-0146386;



PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 61.5%; Score 8; DB 21; Length 183;  
Best Local Similarity 100.0%; Pred. No. 3.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PPPPLPPA 11  
|||||||  
Db 167 PPPPLPPA 174

RESULT 9

AAG07314  
ID AAG07314 standard; Protein; 188 AA.

XX AAG07314;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 4421.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

OS EPI033405-A2.

PN 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156456.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160747.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 61.5%; Score 8; DB 21; Length 188;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||||  
Db 172 PPPPLPPA 179

RESULT 10  
AAG07313  
ID AAG07313 standard; Protein: 195 AA.  
XX  
AC AAG07313;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 4420.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135153.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138647.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 24-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142820.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 23-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 23-AUG-1999; 99US-015066.  
 PR 25-AUG-1999; 99US-0150884.  
 PR 26-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151338.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 28-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157717.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.

PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 61.5%; Score 8; DB 21; Length 195;  
Best Local Similarity 100.0%; Pred. No. 3.4;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPPA 11  
|||||||  
DB 179 PPPPLPPA 186

RESULT 11  
AAB32786  
ID AAB32786 standard; Protein: 234 AA.

XX AC AAB32786;

XX DT 25-JAN-2001 (first entry)

XX DE Eucalyptus grandis transcription factor protein sequence #244.

XX KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
KW poplar; sweetgum; teak; mangany; bZIP; G-box binding factor;  
KW basic helix-loop-helix zipper; homeotic; homeobox; MADS;  
KW homeodomain zipper; LIM domain; AP2; EREBS; zinc finger domain;  
KW type 2 Cys2His2; CCAAT box element; MYB.

XX OS Eucalyptus grandis.

XX PN WO200053724-A2.

XX PD 14-SEP-2000.

XX PF 09-MAR-2000; 2000WO-US06112.

XX PR 11-MAR-1999; 99US-0266513.

XX PR 18-AUG-1999; 99US-0149485.

XX PA (GENE-) GENESIS RES & DEV CORP LTD.  
(FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX PI Wood M, McGrath A, Shenk MA, Glenn M;

XX DR WPI; 2000-579369/54.

XX PT New isolated polynucleotide encoding a plant transcription factor for  
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
PT having modified gene expression or modified activity of a polypeptide

XX PS Claim 8; Page 304; 747pp; English.

XX CC The present invention relates to novel plant transcription factors from  
CC Eucalyptus grandis or Pinus radiata. The present sequence is one such

CC transcription factor. The transcription factor may be used to produce a  
CC plant having modified gene expression such as a woody plant e.g. a  
CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mangany species or  
CC to modify the activity of a polypeptide in a plant. The transcription  
CC factors of the present invention are members from the following families  
CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic  
CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeobox  
CC zipper, LIM domain, AP2 and EREBS, zinc finger domains of type 2  
CC Cys2His2, CCAAT box elements and MYB.

SO Sequence 234 AA;

Query Match 61.5%; Score 8; DB 21; Length 234;  
Best Local Similarity 100.0%; Pred. No. 4;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPPA 11  
|||||||  
DB 35 PPPPLPPA 42

RESULT 12

ID AAB79117 standard; Protein: 313 AA.

XX AC AAB79117;

XX DT 30-APR-2001 (first entry)

XX DE Corynebacterium glutamicum HA protein sequence SEQ ID NO:190.

XX KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;  
KW fine chemical production; organic acid; proteinogenic amino acid;  
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
KW diagnosis; Corynebacterium diptheriae; genetic engineering;  
KW Brevibacterium; environmental condition.

XX OS Corynebacterium glutamicum.

XX PN WO200100842-A2.

XX PD 04-JAN-2001.

XX PF 23-JUN-2000; 2000WO-IB00911.

XX PR 25-JUN-1999; 99US-0141031.

XX PR 08-JUL-1999; 99DE-1031636.

XX PR 09-JUL-1999; 99DE-1032125.

XX PR 09-JUL-1999; 99DE-1032126.

XX PR 09-JUL-1999; 99DE-1032127.

XX PR 09-JUL-1999; 99DE-1032128.

XX PR 09-JUL-1999; 99DE-1032129.

XX PR 09-JUL-1999; 99DE-1032226.

XX PR 14-JUL-1999; 99DE-1032920.

XX PR 14-JUL-1999; 99DE-1032922.

XX PR 14-JUL-1999; 99DE-1032924.

XX PR 14-JUL-1999; 99DE-1032928.

XX PR 14-JUL-1999; 99DE-1032930.

XX PR 14-JUL-1999; 99DE-1032933.

XX PR 14-JUL-1999; 99DE-1032935.

XX PR 14-JUL-1999; 99DE-1032973.

XX PR 14-JUL-1999; 99DE-1033002.

XX PR 14-JUL-1999; 99DE-1033003.

XX PR 14-JUL-1999; 99DE-1033005.

XX PR 14-JUL-1999; 99DE-1033006.

XX PR 31-AUG-1999; 99DE-1041378.

XX PR 31-AUG-1999; 99DE-1041379.

XX PR 31-AUG-1999; 99DE-1041380.

XX PR 31-AUG-1999; 99DE-1041391.

XX PR 03-SEP-1999; 99DE-1042088.

XX (BADI ) BASF AG.  
 XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
 PI WPI: 2001-061974/07.  
 XX N-PSDB; AAF71232.  
 DR  
 XX  
 PT New isolated Corynebacterium glutamicum nucleic acid for production or  
 PT modulation of production of fine chemicals such as amino acids,  
 PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins  
 PT or enzymes -  
 PS  
 XX Claim 20; Page 399-400; 712pp; English.  
 XX  
 CC AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis  
 CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The  
 CC C. glutamicum HA genes (I) can be used in vectors for expression in host  
 CC cells and production of fine chemicals, such as, an organic acid,  
 CC proteinogenic or nonproteinogenic amino acid (preferred), putine or  
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated  
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,  
 CC polypeptide or enzyme. The amino acids produced can be lysine, glutamine,  
 CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,  
 CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,  
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can  
 CC be modulated. The presence of (I) or HA proteins encoded by them are  
 CC used for diagnosing the presence or activity of Corynebacterium  
 CC diptheriae. (I) can be used to map the C. glutamicum genome or can be  
 CC used as markers for genetically engineered Corynebacterium or  
 CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain  
 CC homeostasis in C. glutamicum or help the microorganism to adapt to  
 CC different environmental conditions.  
 CC  
 SQ Sequence 313 AA;

Query Match 61.5%; Score 8; DB 22; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 APPPLPP 10  
 |||||  
 Db 290 apppplpp 297

RESULT 13  
 AAB79242  
 ID AAB79242 standard; Protein: 313 AA.  
 XX  
 AC AAB79242;  
 XX  
 DT 30-APR-2001 (first entry)  
 XX  
 DE Corynebacterium glutamicum HA protein sequence SEQ ID NO:440.  
 XX  
 KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; putine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;  
 KW diagnosis; Corynebacterium diptheriae; genetic engineering;  
 KW Brevibacterium; environmental condition.  
 KW  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200100842-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 23-JUN-2000; 2000MO-IB00911.  
 XX  
 PR 25-JUN-1999; 99US-0141031.

PR 08-JUL-1999; 99DE-1031636.  
 PR 09-JUL-1999; 99DE-1032125.  
 PR 09-JUL-1999; 99DE-1032126.  
 PR 09-JUL-1999; 99DE-1032127.  
 PR 09-JUL-1999; 99DE-1032128.  
 PR 09-JUL-1999; 99DE-1032129.  
 PR 09-JUL-1999; 99DE-1032226.  
 PR 14-JUL-1999; 99DE-1032920.  
 PR 14-JUL-1999; 99DE-1032922.  
 PR 14-JUL-1999; 99DE-1032924.  
 PR 14-JUL-1999; 99DE-1032928.  
 PR 14-JUL-1999; 99DE-1032930.  
 PR 14-JUL-1999; 99DE-1032933.  
 PR 14-JUL-1999; 99DE-1032935.  
 PR 14-JUL-1999; 99DE-1032973.  
 PR 14-JUL-1999; 99DE-1033002.  
 PR 14-JUL-1999; 99DE-1033003.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 31-AUG-1999; 99DE-1041390.  
 PR 31-AUG-1999; 99DE-1041391.  
 PR 03-SEP-1999; 99DE-1042088.  
 XX  
 PA (BADI ) BASF AG.  
 XX  
 PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
 XX  
 DR WPI: 2001-061974/07.  
 XX N-PSDB; AAF71357.  
 XX  
 PT New isolated Corynebacterium glutamicum nucleic acid for production or  
 PT modulation of production of fine chemicals such as amino acids,  
 PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins  
 PT or enzymes -  
 PS  
 XX Claim 20; Page 711-712; 712pp; English.  
 XX  
 CC AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis  
 CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The  
 CC C. glutamicum HA genes (I) can be used in vectors for expression in host  
 CC cells and production of fine chemicals, such as, an organic acid,  
 CC proteinogenic or nonproteinogenic amino acid (preferred), putine or  
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated  
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,  
 CC polypeptide or enzyme. The amino acids produced can be lysine, glutamine,  
 CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,  
 CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,  
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can  
 CC be modulated. The presence of (I) or HA proteins encoded by them are  
 CC used for diagnosing the presence or activity of Corynebacterium  
 CC diptheriae. (I) can be used to map the C. glutamicum genome or can be  
 CC used as markers for genetically engineered Corynebacterium or  
 CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain  
 CC homeostasis in C. glutamicum or help the microorganism to adapt to  
 CC different environmental conditions.  
 CC  
 SQ Sequence 313 AA;

Query Match 61.5%; Score 8; DB 22; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 APPPLPP 10  
 |||||  
 Db 290 apppplpp 297

RESULT 14  
 AAG90541  
 ID AAG90541 standard; Protein: 438 AA.

XX AAG90541;  
 AC 26-SEP-2001 (first entry)  
 DT  
 XX C glutamicum protein fragment SEQ ID NO: 4295.  
 DE  
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 KW  
 XX Corynebacterium glutamicum.  
 OS  
 XX EPI108790-A2.  
 PN  
 XX 20-JUN-2001.  
 PD  
 XX 18-DEC-2000; 2000EP-0127688.  
 PF  
 XX 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 PA (KYOW ) KYOMA HAKKO KOGYO KK.  
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX  
 DR WPI: 2001-376931/40.  
 DR N-PSDB; AAH65760.  
 XX  
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 PS Claim 17; SEQ ID NO: 4295; 246bp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from Coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 CC  
 XX Sequence 438 AA;  
 SQ

Query Match 61.5%; Score 8; DB 22; Length 438;  
 Best Local Similarity 100.0%; Pred. No. 6.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 3 APPPLPP 10  
 DB 314 appplpp 321

RESULT 15  
 AAB93086  
 ID AAB93086 standard; Protein; 586 AA.  
 XX  
 AC AAB93086;  
 XX  
 DT 26-JUN-2001 (first entry)  
 DE  
 XX Human protein sequence SEQ ID NO:11926.  
 DE  
 XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI074617-A2.  
 XX  
 PD 07-FEB-2001.  
 PD  
 XX 28-JUL-2000; 2000EP-0116126.  
 PF  
 XX 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HEIT-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI: 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 PS  
 PS Claim 8; SEQ ID 11926; 2537p + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification, where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03106 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 XX Sequence 586 AA;  
 SQ

Query Match 61.5%; Score 8; DB 22; Length 586;  
 Best Local Similarity 100.0%; Pred. No. 8.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 4 PPPPLPPA 11  
 DB 18 pppplppa 25

Search completed: January 22, 2002, 16:31:44  
 Job time: 324 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:32:17 ; Search time 26.82 Seconds  
(without alignments)  
36.923 Million cell updates/sec

Title: US-09-485-529-103

Perfect score: 13

Sequence: 1 LNAPPPPLPPAPQ 13

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: pir68:\*  
2: pir1:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	61.5	443	2	T29820
2	8	61.5	564	2	T45866
3	8	61.5	907	2	E96636
4	8	61.5	1032	2	D83637
5	8	61.5	1520	2	T00273
6	7	53.8	138	2	C96734
7	7	53.8	145	2	T48552
8	7	53.8	219	2	I51382
9	7	53.8	238	2	T32889
10	7	53.8	253	2	T17312
11	7	53.8	281	2	T38707
12	7	53.8	285	2	T18689
13	7	53.8	289	2	T52354
14	7	53.8	341	2	D96500
15	7	53.8	359	2	T13478
16	7	53.8	373	2	A70856
17	7	53.8	389	2	S27200
18	7	53.8	431	1	E70699
19	7	53.8	433	2	T07910
20	7	53.8	477	2	I38409
21	7	53.8	502	2	A55197
22	7	53.8	538	1	FOVW1M
23	7	53.8	555	2	T03049
24	7	53.8	574	2	T05964
25	7	53.8	585	2	T00979
26	7	53.8	588	2	T45564
27	7	53.8	596	2	T03908
28	7	53.8	656	1	QOBE72
29	7	53.8	664	2	T01368

30	7	53.8	678	2	T04832	probable serine/th
31	7	53.8	708	2	D96711	hypothetical prote
32	7	53.8	716	2	T26598	hypothetical prote
33	7	53.8	736	2	I51691	dishevelled homolo
34	7	53.8	820	2	T46412	ubiquitin--protein
35	7	53.8	933	1	QRHUP	progesterone recep
36	7	53.8	981	1	FOVWGM	gag-abl polyprotel
37	7	53.8	994	2	S19595	chloride channel p
38	7	53.8	998	2	T30930	hypothetical prote
39	7	53.8	1076	2	S50536	hypothetical prote
40	7	53.8	1110	2	T19673	hypothetical prote
41	7	53.8	1145	2	T18235	transcription acti
42	7	53.8	1206	2	S24407	formin isoform IV
43	7	53.8	1264	2	A36858	G2R protein - vari
44	7	53.8	1268	2	T31420	C-terminal domain
45	7	53.8	1344	2	T14316	rig-I protein - mo

#### ALIGNMENTS

RESULT 1  
T29820  
hypothetical protein F25E2.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
R:Minx, P.; Le, T.  
C:Accession: T29820  
submitted to the EMBL Data Library, March 1996  
A:Description: The sequence of C. elegans cosmid F25E2.  
A:Reference number: Z20691  
A:Accession: T29820  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-443 <MIN>  
A:Cross-references: EMBL:U50197; PIDN:AAA91257.1; CESP:F25E2.4  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:F25E2.4  
A:introns: 72/3; 110/3; 143/2; 171/3; 330/3; 385/3

Query Match 61.5%; Score 8; DB 2; Length 443;  
Best Local Similarity 100.0%; Pred.No. 1.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 NAPPPLP 9  
DB 415 NAPPPLP 422

RESULT 2  
T45866  
hypothetical protein F3A4.210 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T45866  
R:Barques, M.; Collado, M.C.; Navarro, P.; Terol, J.; Perez-Alonso, M.; Mewes, H.W.;  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: Z23007  
A:Accession: T45866  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-564 <BAR>  
A:Cross-references: EMBL:AL132978  
A:Experimental source: cultivar Columbia; BAC clone F3A4  
C:Genetics:  
A:Map position: 3  
A:introns: 250/2; 424/1  
A:Note: F3A4.210

Query Match 61.5%; Score 8; DB 2; Length 564;

Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||  
DB 51 PPPPLPPA 58

## RESULT 3

E96636  
hypothetical protein T7P1.21 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: E96636

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719

A:Accession: E96636

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-907 <STO>  
A:Cross-references: GB:AE005173; NID:96751696; PIDN:AAE27679.1; GSPDB:GN00141

C:Genetics:

A:Gene: T7P1.21

A:Map position: 1

Query Match  
Best Local Similarity 100.0%; Pred. No. 2.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||  
DB 476 PPPPLPPA 483

## RESULT 4

DB3637  
serine/threonine protein kinase PpxA PA0074 [imported] - Pseudomonas aeruginosa (strain

C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: DB3637

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Marrener, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim, ; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337

A:Accession: DB3637

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1032 <STO>

A:Cross-references: GB:AE004446; GB:AE004091; NID:9945886; PIDN:AA03464.1; GSPDB:GN001

C:Genetics:

A:Experimental source: strain PA01

A:Gene: ppxA; PA0074

Query Match  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||  
DB 337 PPPPLPPA 344

RESULT 5  
T00273  
hypothetical protein KIAA0595 - human (fragment)

C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000

C:Accession: T00273

R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, DNA Res. 5, 31-39, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp

A:Reference number: Z14086; MUID:98290545

A:Accession: T00273

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1520 <NAG>

A:Cross-references: EMBL:AB011167; NID:93043713; PIDN:BA25521.1; PID:93043714

C:Genetics:

A>Note: KIAA0595

Query Match  
Best Local Similarity 100.0%; Pred. No. 4.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||  
DB 854 PPPPLPPA 861

## RESULT 6

C96734  
hypothetical protein F15H11.18 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: C96734

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719

A:Accession: C96734

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-138 <STO>

A:Cross-references: GB:AE005173; NID:95902409; PIDN:AAD5511.1; GSPDB:GN00141

C:Genetics:

A:Gene: F15H11.18

A:Map position: 1

Query Match  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10  
|||||  
DB 71 PPPPLPP 77

## RESULT 7

T48552  
glutaredoxin-like protein - Arabidopsis thaliana

N:Alternate names: protein F14F18.100  
C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48552  
R:Byvan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, April 2000  
A:Reference number: 224490  
A:Accession: T48552  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-145 <BEV>  
A:Cross-references: EMBL:AL163812  
A:Experimental source: cultivar Columbia; BAC clone F14F18  
C:Genetics:  
A:Map position: 5  
A:Note: F14F18.100

Query Match 53.8%; Score 7; DB 2; Length 145;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
|||||  
DB 22 PPPPLPP 28

RESULT 8  
151382  
achaete-scute homolog - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: 151382  
R:Jasoni, C.L.; Walker, M.B.; Morris, M.D.; Reh, T.A.  
Development 120, 769-783, 1994  
A:Title: A chicken achaete-scute homolog (CASH-1) is expressed in a temporally and spatially  
A:Reference number: 151382; MUID:95324365  
A:Accession: 151382  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-219 <JAS>  
A:Cross-references: EMBL:U01339; NID:g401726; PIDN:AAC59658.1; PID:g401727

Query Match 53.8%; Score 7; DB 2; Length 219;  
Best Local Similarity 100.0%; Pred. No. 6.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 APPPLP 9  
|||||  
DB 42 APPPLP 48

RESULT 9  
T32889  
hypothetical protein C34B2.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T32889  
R:Graves, T.; Sutterer, C.; Hawkins, M.; Wilson, R.  
Submitted to the EMBL Data Library, January 1998  
A:Description: The sequence of C. elegans cosmid C34B2.  
A:Reference number: 221241  
A:Accession: T32889  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-238 <GRA>  
A:Cross-references: EMBL:AF043693; PIDN:AAB97541.1; GSPDB:GN00019; CESP:C34B2.9  
A:Experimental source: strain Bristol N2; clone C34B2  
C:Genetics:  
A:Gene: CESP:C34B2.9  
A:Map position: 1  
A:introns: 35/3; 91/2; 200/2

Query Match 53.8%; Score 7; DB 2; Length 238;

Best Local Similarity 100.0%; Pred. No. 6.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
|||||  
DB 132 PPPPLPP 138

RESULT 10  
T17312  
hypothetical protein DKFZp434H018.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T17312  
R:Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, September 1999  
A:Reference number: Z18726  
A:Accession: T17312  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-253 <OT>  
A:Cross-references: EMBL:AL117579  
A:Experimental source: adult testis; clone DKFZp434H018  
C:Genetics:  
A:Note: DKFZp434H018.1

Query Match 53.8%; Score 7; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
|||||  
DB 35 PPPPLPP 41

RESULT 11  
I38707  
Fas ligand - human  
C:Species: Homo sapiens (man)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jul-2000  
C:Accession: I38707; JC2340; S57565; I38554  
R:Takahashi, T.; Tanaka, M.; Inazawa, J.; Abe, T.; Suda, T.; Nagata, S.  
Int. Immunol. 6, 1567-1574, 1994  
A:Title: Human Fas ligand: gene structure, chromosomal location and species specificity  
A:Reference number: I38707; MUID:95127560  
A:Accession: I38707  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-281 <RES>  
A:Cross-references: EMBL:U11821; NID:g595430; PIDN:AAC50124.1; PID:g595431  
R:Mita, E.; Hayashi, N.; Iio, S.; Takehara, T.; Hijioka, T.; Kasahara, A.; Fusamoto, Blochem. Biophys. Res. Commun. 204, 468-474, 1994  
A:Title: Role of Fas ligand in apoptosis induced by hepatitis C virus infection.  
A:Reference number: JC2340; MUID:95071350  
A:Accession: JC2340  
A:Molecule type: DNA  
A:Residues: 1-281 <MIT>  
A:Cross-references: GB:D38122; DDBJ:D29820; NID:g601892; PIDN:BAA07320.1; PID:g136990  
R:Schatzlein, C.E.  
submitted to the EMBL Data Library, June 1995  
A:Reference number: S57565  
A:Accession: S57565  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-281 <SCH>  
A:Cross-references: EMBL:X89102; NID:g887455; PID:g887456  
R:Alderson, M.R.; Tough, T.W.; Davis-Smith, I.; Braddy, S.; Falk, B.; Schooley, K.A.; J. Exp. Med. 181, 71-77, 1995  
A:Title: Fas ligand mediates activation-induced cell death in human T lymphocytes.  
A:Reference number: I38554; MUID:95105731  
A:Accession: I38554  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
A:Residues: 1-281 <RE2>  
A:Cross-references: EMBL:U08137; NID:g624627; PIDN:AAC50071.1; PID:g624628  
C:Genetics:  
A:Gene: FasI  
A:Introns: 151/1; 116/3  
C:Keywords: glycoprotein; transmembrane protein  
F:80-102/Domain: transmembrane #status predicted  
F:76,184,250,260/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 53.8%; Score 7; DB 2; Length 281;  
Best Local Similarity 100.0%; Pred. No. 8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 PPPPLPP 10  
|||  
Db 50 PPPPLPP 56

## RESULT 12

hypothetical protein B0285.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18689  
R:Stinson, J.  
submitted to the EMBL Data Library, June 1994  
A:Reference number: Z19007  
A:Accession: T18689  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-285 <MTL>  
A:Cross-references: EMBL:Z34533; PIDN:CAA84294.1; GSPDB:GN00021; CESP:B0285.2  
A:Experimental source: clone B0285  
C:Genetics:  
A:Gene: CESP:B0285.2  
A:Map position: 3  
A:Introns: 65/2; 131/2; 173/3; 211/3

Query Match 53.8%; Score 7; DB 2; Length 285;  
Best Local Similarity 100.0%; Pred. No. 8.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 PPPPLPP 10  
|||  
Db 207 PPPPLPP 213

## RESULT 13

hypothetical protein B11E6.30 [imported] - Neurospora crassa  
C:Species: Neurospora crassa  
C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000  
C:Accession: T52354  
R:Schulte, U.; Aitn, V.; Hohnel, J.; Brandt, P.; Farman, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, September 2000  
A:Reference number: Z26053  
A:Accession: T52354  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-289 <SCH>  
A:Cross-references: EMBL:AL442043; GSPDB:GN00116; NCSP:B11E6.30  
A:Experimental source: BAC clone B11E6; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B11E6.30  
A:Map position: 6

Query Match 53.8%; Score 7; DB 2; Length 289;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 PPPPLPP 10  
|||  
Db 43 PPPPLPP 49

## RESULT 14

probable transcription factor [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: D96500  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hultzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719  
A:Accession: D96500  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-341 <STO>  
A:Cross-references: GB:AE005173; NID:g7523681; PIDN:AAE63120.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F2J6.6  
A:Map position: 1

Query Match 53.8%; Score 7; DB 2; Length 341;  
Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LNAPPPP 7  
|||  
Db 63 LNAPPPP 69

## RESULT 15

hypothetical protein 34F3.10 - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: T13478  
R:Valenti, P.; Salles, C.; Campbell, L.; Glover, D.  
submitted to the EMBL Data Library, April 1999  
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A:Reference number: Z17685  
A:Accession: T13478  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-359 <PHI>  
A:Cross-references: EMBL:AL031583; NID:e1321005; PID:e1321018; PIDN:CA841346.1  
C:Genetics:  
A:Cross-references: FlyBase:FBgn0025623  
A:Introns: 17/2; 50/3; 333/2  
A:Note: EG:34F3.10

Query Match 53.8%; Score 7; DB 2; Length 359;  
Best Local Similarity 100.0%; Pred. No. 9.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 PPPPLPP 10  
|||  
Db 170 PPPPLPP 176

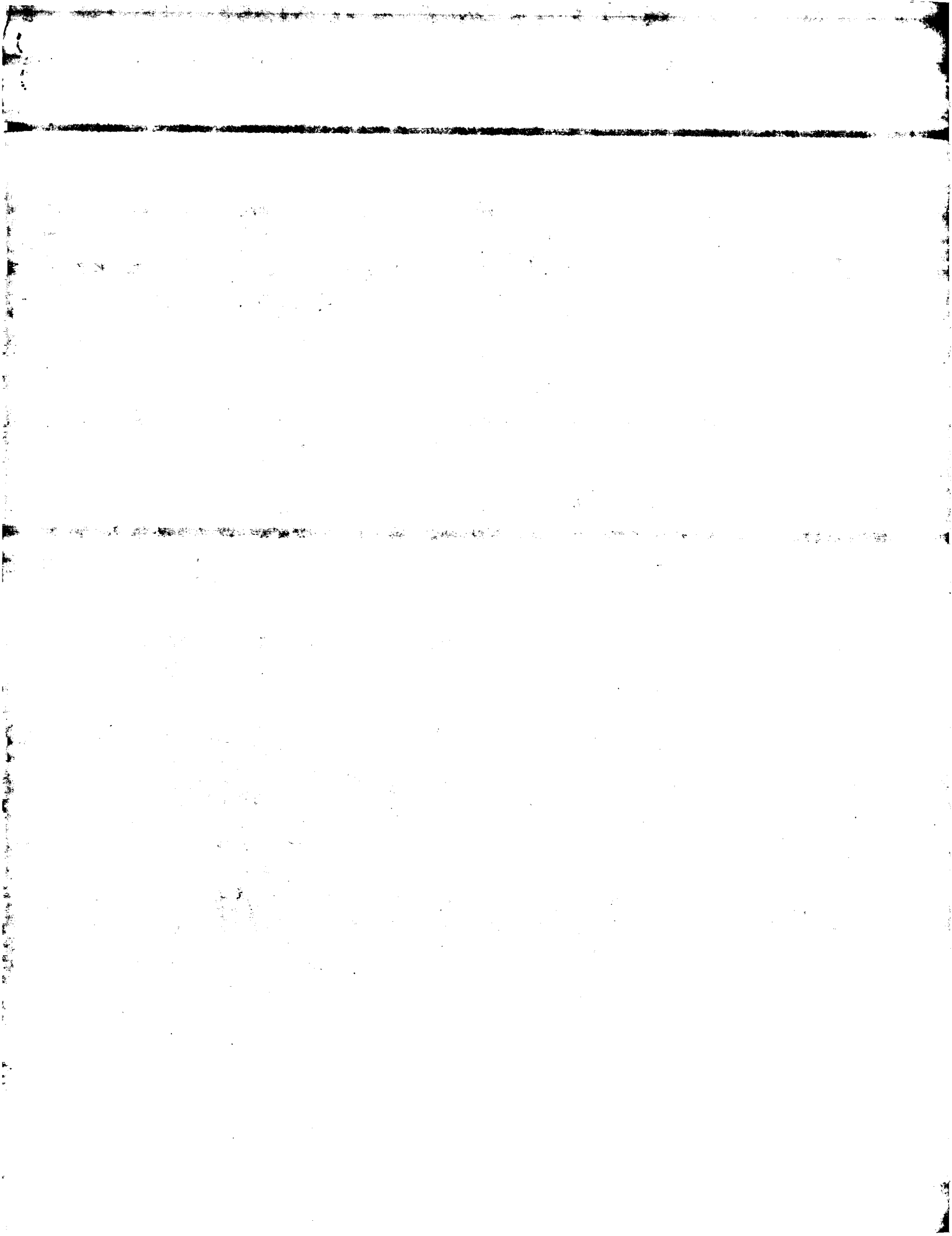
Thu Jan 24 09:56:01 2002

us-09-485-529-103.oli.rpr

Page 5

Search completed: January 22, 2002, 16:32:18  
Job time: 178 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:37:08 ; Search time 17.09 Seconds  
(without alignments)  
27.890 Million cell updates/sec

Title: US-09-485-529-103  
Perfect score: 13  
Sequence: 1 LNAPPPLPPAPQ 13

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	69.2	282	1	ATF5_HUMAN
2	8	61.5	591	1	MNT_MOUSE
3	7	53.8	235	1	GAG_MLYAB
4	7	53.8	281	1	PAGL_HUMAN
5	7	53.8	285	1	YK2_CAEEL
6	7	53.8	288	1	SMN_MOUSE
7	7	53.8	290	1	TRX2_MOUSE
8	7	53.8	314	1	Y009_HUMAN
9	7	53.8	331	1	PELB_COLGL
10	7	53.8	340	1	GBX2_CHICK
11	7	53.8	389	1	NDP_MOUSE
12	7	53.8	408	1	G3PT_MOUSE
13	7	53.8	412	1	ALF_PETHY
14	7	53.8	431	1	PKNA_MYCTU
15	7	53.8	453	1	GAAG_HUMAN
16	7	53.8	477	1	CARP2_HUMAN
17	7	53.8	502	1	WASP_HUMAN
18	7	53.8	538	1	GAG_MLYMO
19	7	53.8	559	1	WAS1_HUMAN
20	7	53.8	656	1	UL25_HCMVA
21	7	53.8	684	1	CDK9_CAEEL
22	7	53.8	724	1	CC11_MOUSE
23	7	53.8	726	1	CC11_HUMAN
24	7	53.8	727	1	CC11_HORSE
25	7	53.8	736	1	DVL2_XENLA
26	7	53.8	757	1	CC11_MOUSE
27	7	53.8	757	1	CC11_MOUSE
28	7	53.8	809	1	CNAD_HUMAN
29	7	53.8	884	1	ANDR_EULFC
30	7	53.8	901	1	Y298_HUMAN
31	7	53.8	933	1	PRGR_HUMAN
32	7	53.8	971	1	Y029_HUMAN
33	7	53.8	994	1	CLC1_RAT

## ALIGNMENTS

RESULT	ID	Score	Query Match	Length	ID	Description
1	ATF5_HUMAN	9	69.2	282	1	ATF5_HUMAN
2	ATF5_HUMAN	8	61.5	591	1	MNT_MOUSE
3	ATF5_HUMAN	7	53.8	235	1	GAG_MLYAB
4	ATF5_HUMAN	7	53.8	281	1	PAGL_HUMAN
5	ATF5_HUMAN	7	53.8	285	1	YK2_CAEEL
6	ATF5_HUMAN	7	53.8	288	1	SMN_MOUSE
7	ATF5_HUMAN	7	53.8	290	1	TRX2_MOUSE
8	ATF5_HUMAN	7	53.8	314	1	Y009_HUMAN
9	ATF5_HUMAN	7	53.8	331	1	PELB_COLGL
10	ATF5_HUMAN	7	53.8	340	1	GBX2_CHICK
11	ATF5_HUMAN	7	53.8	389	1	NDP_MOUSE
12	ATF5_HUMAN	7	53.8	408	1	G3PT_MOUSE
13	ATF5_HUMAN	7	53.8	412	1	ALF_PETHY
14	ATF5_HUMAN	7	53.8	431	1	PKNA_MYCTU
15	ATF5_HUMAN	7	53.8	453	1	GAAG_HUMAN
16	ATF5_HUMAN	7	53.8	477	1	CARP2_HUMAN
17	ATF5_HUMAN	7	53.8	502	1	WASP_HUMAN
18	ATF5_HUMAN	7	53.8	538	1	GAG_MLYMO
19	ATF5_HUMAN	7	53.8	559	1	WAS1_HUMAN
20	ATF5_HUMAN	7	53.8	656	1	UL25_HCMVA
21	ATF5_HUMAN	7	53.8	684	1	CDK9_CAEEL
22	ATF5_HUMAN	7	53.8	724	1	CC11_MOUSE
23	ATF5_HUMAN	7	53.8	726	1	CC11_HUMAN
24	ATF5_HUMAN	7	53.8	727	1	CC11_HORSE
25	ATF5_HUMAN	7	53.8	736	1	DVL2_XENLA
26	ATF5_HUMAN	7	53.8	757	1	CC11_MOUSE
27	ATF5_HUMAN	7	53.8	757	1	CC11_MOUSE
28	ATF5_HUMAN	7	53.8	809	1	CNAD_HUMAN
29	ATF5_HUMAN	7	53.8	884	1	ANDR_EULFC
30	ATF5_HUMAN	7	53.8	901	1	Y298_HUMAN
31	ATF5_HUMAN	7	53.8	933	1	PRGR_HUMAN
32	ATF5_HUMAN	7	53.8	971	1	Y029_HUMAN
33	ATF5_HUMAN	7	53.8	994	1	CLC1_RAT

DR	PRINTS; PR01217; PRICHTEXTENS.
DR	PDomon; PD004952; HamP-like; 1.
DR	SMART; SM00338; BR_L; 1.
DR	PROSITE; PS00036; BZIP_BASIC; 1.
KW	Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW	Multigene family.
FT	DOMAIN 123 139 POLY-PRO.
FT	DOMAIN 186 194 POLY-PRO.
FT	DNA_BIND 210 230 BASIC MOTIF.
FT	DOMAIN 236 250 LEUCINE-ZIPPER (PROBABLE).
FT	CONFLICT 161 163 LLA -> RHE (IN REF. 3)
SQ	SEQUENCE 282 AA; 30674 MW; DBB2F907CA0215A0 CRC64;
Query Match	
Best Local Similarity	69.2%; Score 9; DB 1; Length 282;
Matches	9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy	4 PEPPLPPAP 12
db	 131 PEPPLPPAP 139

	RESULT	2
ID	MNT_MOUSE	
AC	NMT_MOUSE	STANDARD;
AD	O08789; P97349;	PRT: 591 AA.
DT	15-DEC-1998 (Rel. 37, Created)	
DT	15-DEC-1998 (Rel. 37, Last sequence update)	
DT	15-DEC-1998 (Rel. 37, Last annotation update)	
DE	MAX BINDING PROTEIN MNT (ROX PROTEIN) (MYC ANTAGONIST MNT).	
CN	NMT OR ROX.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_Taxid=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
KC	TISSUE=Embryo;	
KX	MEDLINE=97152466; PubMed=9000049;	
RA	Hurtin P.J., Queva C., Eisenman R.N.;	
RT	"Mnt, a novel Max-interacting protein is coexpressed with Myc in	
RL	proliferating cells and mediates repression at Myc binding sites.";	
RN	Genes Dev. 11:44-56(1997).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
KC	TISSUE=Embryo;	
KX	MEDLINE=97327566; PubMed=9184233;	
RA	Meroni G., Raymond A., Alcalay M., Borsani G., Taniguchi A.,	
RA	Tenoloreni R., Lo Negro C., Messali S., Zollo M., Ledbetter D.H.,	
RA	Brent R., Ballabio A., Carrozzo R.;	
RT	"Rox, a novel bHLH/p protein expressed in quiescent cells that	
RT	heterodimerizes with Max, binds a non-canonical E box and acts as a	
RT	transcriptional repressor.";	
RL	EMBO J. 16:2892-2906(1997).	
CC	-1- FUNCTION: BINDS DNA AS A HETERODIMER WITH MAX AND REPRESSES	
CC	TRANSCRIPTION. BINDS TO THE CANONICAL E BOX SEQUENCE 5'-CACGTG-3'	
CC	AND, WITH HIGHER AFFINITY, TO 5'-CAAGCG-3'.	
CC	-1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER	
CC	BHLH PROTEIN. BINDS DNA AS AN HOMODIMER OR A HETERODIMER WITH MAX.	
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.	
CC	-1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF	
CC	TRANSCRIPTION FACTORS. BHLH-ZIP SUPERFAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on ways	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ,	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
OR	EMBL; U77356; AACB38687.1; -.	

SEQUENCE	591 AA;	63311 MM;	34EB61A10CF4142F	CNC64;
DR EMBL; Y07609; CAA68878.1; -.				
DR MGD; MGI:109150; Mnt.				
DR InterPro; IPR003015; HLH_MYC.				
DR InterPro; IPR001092; HLH_dlm.				
DR Pfam; Pf00010; HLH; 1.				
DR SMART; SM00353; HLH; 1.				
DR PROSITE; PS00036; HELIX_LOOP_HELIX; FALSE_NEG.				
KW Transcription regulation; Repressor; Nuclear protein; DNA-binding.				
FT DNA_BIND	224	235		
FT DOMAIN	236	272		
FT DOMAIN	273	301		
FT CONFLICT	379	379		
FT CONFLICT	392	392		
FT CONFLICT	402	403		
FT CONFLICT	414	414		
FT CONFLICT	431	431		
FT CONFLICT	465	465		
FT CONFLICT	525	525		
FT CONFLICT	558	558		
FT CONFLICT	558	558		
SEQUENCE	591 AA;	63311 MM;	34EB61A10CF4142F	CNC64;

[illegible]



```

FT CHAIN 2 131 CORE PROTEIN P15.
FT CHAIN 132 215 INNER COAT PROTEIN P12.
FT CHAIN 216 235 CORE SHELL PROTEIN P30.
FT LIPID 2 2 MYRISTATE.
SQ SEQUENCE 235 AA; 25641 MW; 4DB3F71D7E056C7D CRC64;

Query Match 53.8%; Score 7; DB 1; Length 235;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10
DB 105 PPPPLPP 111

RESULT 4
FASL_HUMAN
ID FASL_HUMAN STANDARD; PRT; 281 AA.
AC P48023;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FAS ANTIGEN LIGAND (APOPTOSIS ANTIGEN LIGAND) (APTL).
GN TNFSF6 OR APTL1G1 OR FASL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=95105731; PubMed=7528780;
RA Alderson M.;
RT "Fas ligand mediates activation-induced cell death in human T
RT lymphocytes.";
RL J. Exp. Med. 181:71-77(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95127560; PubMed=7826947;
RA Takahashi T., Tanaka M., Inazawa J., Abe T., Suda T., Nagata S.;
RT "Human Fas ligand: gene structure, chromosomal location and species
RT specificity.";
RL Int. Immunol. 6:1567-1574(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Schaeuble C.E., Philippson P., Eibel H.;
RT Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=95071350; PubMed=7980502;
RA Mita E., Hayashi N., Ito S., Takehana T., Hijioka T., Kasahara A.,
RA Fusamoto H., Kamada T.;
RT "Role of Fas ligand in apoptosis induced by hepatitis C virus
RT infection.";
RL Biochem. Biophys. Res. Commun. 204:468-474(1994).
RN [5]
RP SEQUENCE FROM N.A.
RA Wilkinson J.;
RT Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.
RN [6]
RP SEQUENCE OF 1-10 FROM N.A.
RC TISSUE=Blood;
RA Matsumura M., Nakanishi Y., Ohba Y.;
RT Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CYTOKINE THAT BINDS TO FAS ANTIGEN, A RECEPTOR THAT
CC TRANSDUCES THE APOPTOTIC SIGNAL INTO CELLS. MAY BE INVOLVED IN
CC CYTOTOXIC T CELL MEDIATED APOPTOSIS AND IN T CELL DEVELOPMENT.
CC FAS ANTIGEN MEDIATED APOPTOSIS MAY HAVE A ROLE IN THE INDUCTION OF
CC PERIPHERAL TOLERANCE. IN THE ANTIGEN-STIMULATED SUICIDE OF MATURE
CC T CELLS, OR BOTH.
CC -1- SUBUNIT: HOMOTRIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MAY BE RELEASED
CC INTO THE EXTRACELLULAR FLUID, PROBABLY BY CLEAVAGE FROM THE CELL

```

```

CC SURFACE
CC -1- SIMILARITY: BELONGS TO THE TUMOR NECROSIS FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X89102; CAA61474.1; -
DR EMBL: U08137; AAC50071.1; -
DR EMBL: U11821; AAC50124.1; -
DR EMBL: D38122; BAA07320.1; -
DR EMBL: Z96050; CAB09424.1; -
DR EMBL: AB013303; BAA32542.1; -
DR HSSP: P01375; 2TUN.
DR MIM: 134638; -
DR InterPro: IPR003263; TNF_5.
DR InterPro: IPR000478; TNF_family.
DR Pfam: PF00229; TNF_1.
DR ProDom: PD008600; TNF_5; 1.
DR SMART: SM00207; TNF_1.
DR PROSITE: PS00251; TNF_1; 1.
DR PROSITE: PS50049; TNF_2; 1.
KW Cytokine; Transmembrane; Glycoprotein; Signal-anchor; Apoptosis.
FT DOMAIN 1 80 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 103 102 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 4 70 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 45 65 PRO-RICH.
FT DISULFID 202 233 BY SIMILARITY.
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 281 AA; 31485 MW; A8A6B358246E9BB CRC64;

Query Match 53.8%; Score 7; DB 1; Length 281;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10
DB 50 PPPPLPP 56

RESULT 5
YK62_CAEEL
ID YK62_CAEEL STANDARD; PRT; 285 AA.
AC P46552;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHEICAL 31.4 KDA PROTEIN B0285.2 IN CHROMOSOME III.
GN B0285.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Sulston J.;
RT Submitted (JUN-1994) to the EMBL/Genbank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch).  
 -----  
 CC EMBL; Z34533; CAA84294.1; -  
 CC WormPep; B0285.2; CE00641.  
 KW Hypothetical protein.  
 FT DOMAIN 70 75 POLY-GLY.  
 FT DOMAIN 123 129 POLY-SER.  
 FT DOMAIN 163 170 POLY-SER.  
 FT DOMAIN 189 192 POLY-PRO.  
 FT DOMAIN 206 213 POLY-PRO.  
 FT DOMAIN 252 256 POLY-PRO.  
 SQ SEQUENCE 285 AA; 31396 MW; BFE68FF2038A3337 CRC64;  
 -----  
 Query Match 53.8%; Score 7; DB 1; Length 285;  
 Best Local Similarity 100.0%; Pred. No. 4,7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 PPPPLP 10  
 DB 207 PPPPLP 213  
 -----  
 RESULT 6  
 SMN\_MOUSE STANDARD; PRT; 288 AA.  
 ID SMN\_MOUSE  
 AC P97801; C09092;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SURVIVAL MOTOR NEURON PROTEIN.  
 GN SMN1 OR SMN.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9724505; PubMed=9070939;  
 RA Viollet L., Bertrand S., Brunialti A.L.B., Lefebvre S., Burtet P.,  
 RA Clermont O., Cruaud C., Guenet J.-L., Munnich A., Melki J.,  
 RT "cDNA isolation, expression, and chromosomal localization of the  
 RT mouse survival motor neuron gene (Smn).";  
 RL Genomics 40:185-188(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97264340; PubMed=9110173;  
 RA Diconato C.J., Chen X.N., Noya D., Korenberg J.R., Nadeau J.H.,  
 RA Simard L.R.;  
 RT "Cloning, characterization, and copy number of the murine survival  
 RT motor neuron gene: homolog of the spinal muscular atrophy-determining  
 RT gene.";  
 RL Genome Res. 7:339-352(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-BALB/C; TISSUE=Brain;  
 RX MEDLINE=97420785; PubMed=9275227;  
 RA Schrank B., Goetz R., Gunnarsen J.M., Ure J.M., Toyka K.V.,  
 RA Smith A.G., Sendtner M.;  
 RT "Inactivation of the survival motor neuron gene, a candidate gene for  
 RT human spinal muscular atrophy, leads to massive cell death in early  
 RT mouse embryos.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:9920-9925(1997).  
 CC -1- FUNCTION: THE SMN COMPLEX PLAYS AN ESSENTIAL ROLE IN SPLICESOMAL  
 CC SNRP ASSEMBLY IN THE CYTOPLASM, AND IS REQUIRED FOR PRE-MRNA  
 CC SPLICING IN THE NUCLEUS (BY SIMILARITY).  
 CC -1- SUBUNIT: FORMS A STABLE HETEROMERIC COMPLEX WITH SURVIVAL OF MOTOR  
 CC NEURON PROTEIN INTERACTING PROTEIN 1 (SIP1/GEMIN2), GEMIN3 AND  
 CC GEMIN4. THIS COMPLEX IS THOUGHT TO PLAY AN ESSENTIAL ROLE IN  
 CC BIOGENESIS OF THE SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLES  
 CC (SNRNPs). INTERACTS WITH FIBRILLARIN AND WITH SEVERAL SPLICESOMAL  
 CC SNRP CORE SM PROTEINS, INCLUDING B/B', SM D AND SM E PROTEINS (BY  
 CC SIMILARITY).

-1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR. LOCALIZED IN  
 CC SUBNUCLEAR STRUCTURES NEXT TO COILED BODIES, CALLED GEMs, WHICH  
 CC ARE HIGHLY ENRICHED IN SPLICESOMAL SNRNPs (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SMN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.lsb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC -----  
 CC EMBL; U63294; AAC53057.1; -  
 CC EMBL; U77714; AAC53144.1; -  
 CC EMBL; Y12835; CAA73356.1; -  
 CC MGD; MGI:109257; Smn.  
 DR InterPro; IPR002999; Tudor.  
 DR SMART; SM00333; Tudor; 1.  
 KW mRNA processing; RNA-binding; Nuclear protein.  
 FT DOMAIN 190 196 POLY-PRO.  
 FT DOMAIN 212 222 POLY-PRO.  
 FT DOMAIN 239 243 POLY-PRO.  
 SQ SEQUENCE 288 AA; 31254 MW; 757B3074649F7458 CRC64;  
 -----  
 Query Match 53.8%; Score 7; DB 1; Length 288;  
 Best Local Similarity 100.0%; Pred. No. 4,7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 PPPPLP 10  
 DB 214 PPPPLP 220  
 -----  
 RESULT 7  
 TRX2\_MOUSE STANDARD; PRT; 290 AA.  
 ID TRX2\_MOUSE  
 AC O08550;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE TRITHORAX HOMOLOG 2 (MW DOMAIN BINDING PROTEIN 7) (FRAGMENT).  
 GN TRX2 OR WBP7.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97315177; PubMed=9171351;  
 RA Bedford M.T., Chan D.C., Leder P.;  
 RT "RBP W domains and the Abi SH3 domain bind to a specific class of  
 RT proline-rich ligands.";  
 RL EMBO J. 16:2376-2383(1997).  
 CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION  
 CC FACTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.lsb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC -----  
 CC EMBL; U92455; AAC53192.1; -  
 CC MGD; MGI:109565; Wbp7.  
 DR InterPro; IPR000637; AT\_hook.  
 DR Pfam; PF02178; AT\_hook; 1.  
 DR SMART; SM00384; AT\_hook; 1.

KM DNA-binding: Nuclear protein; Transcription regulation.  
 FT NON\_TER 1 1  
 FT DNA\_BIND <1 7 A.T HOOK (BY SIMILARITY).  
 FT DOMAIN 4 35 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 45 77 POLY-PRO.  
 FT DOMAIN 253 263 POLY-PRO.  
 FT NON\_TER 290 290  
 SQ SEQUENCE 290 AA; 31202 MW; C49B5565FC8D2DF CRC64;  
 NCBI\_TaxID=5457;  
 Query Match 53.8%; Score 7; DB 1; Length 290;  
 Best Local Similarity 100.0%; Pred. No. 4.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 PPPPLP 10  
 Db 50 PPPPLP 56  
 RESULT 8  
 ID Y009\_HUMAN STANDARD; PRT; 314 AA.  
 AC Q15390;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL PROTEIN KIAA0009.  
 GN KIAA0009.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bone marrow;  
 RX MEDLINE=96051387; PubMed=7584026;  
 RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,  
 RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;  
 RT "Prediction of the coding sequences of unidentified human genes. I.  
 RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by  
 RT analysis of randomly sampled cDNA clones from human immature myeloid  
 RT cell line KG-1."  
 RL DNA Res. 1:27-35(1994).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D13634; BAA02798.1; -;  
 KM Hypothetical protein  
 FT DOMAIN 165 178 POLY-PRO.  
 SQ SEQUENCE 314 AA; 34756 MW; E79E4FE31061953 CRC64;  
 Query Match 53.8%; Score 7; DB 1; Length 314;  
 Best Local Similarity 100.0%; Pred. No. 5.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 PPPPLP 10  
 Db 171 PPPPLP 177  
 RESULT 9  
 ID Y009\_HUMAN STANDARD; PRT; 331 AA.  
 AC Q15390;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PEPTATE LYASE B PRECURSOR (EC 4.2.2.2).  
 GN PEB.  
 OS Colletotrichum gloeosporioides (Anthracnose fungus) (Glomerella  
 OS cingulata).  
 OS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;  
 OC Glomerella.  
 OX NCBI\_TaxID=5457;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CG-14;  
 RA Mettard C., Keen N.T.;  
 RU Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ACTS AS A VIRULENCE FACTOR ACTIVE IN PLANT TISSUE  
 CC MACERATION.  
 CC -1- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PEPTATE TO GIVE  
 CC OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-MANN-4-ENURONOSYL GROUPS  
 CC AT THEIR NON-REDUCING ENDS.  
 CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF052632; AAD09857.1; -;  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 DR KW Lyase; Signal.  
 FT SIGNAL 1 2 POTENTIAL.  
 FT CHAIN 99 99 PEPTATE LYASE B.  
 FT CARBOHYD 99 99 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SQ SEQUENCE 331 AA; 35108 MW; 475A76F25D4689 CRC64;  
 Query Match 53.8%; Score 7; DB 1; Length 331;  
 Best Local Similarity 100.0%; Pred. No. 5.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 APPPLP 9  
 Db 18 APPPLP 24  
 RESULT 10  
 ID GBX2\_CHICK STANDARD; PRT; 340 AA.  
 AC O42230;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE HOMEOBOX PROTEIN GBX-2 (GASTRULATION AND BRAIN-SPECIFIC HOMEOBOX  
 DE PROTEIN 2).  
 GN GBX2.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98004222; PubMed=9346236;  
 RA Kowenz-Leutz E., Herr P., Niss K., Leutz A.;  
 RT "The homeobox gene GBX2, a target of the myb oncogene, mediates  
 RT autocrine growth and monocytic differentiation.";  
 RL Cell 91:185-195(1997).  
 CC -1- FUNCTION: MAY ACT AS A TRANSCRIPTION FACTOR. FOR CELL PLURIPOTENCY  
 CC AND DIFFERENTIATION IN THE EMBRYO.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF02151; AAB82710.1; -;  
 CC InterPro: IPR01356; Homeobox.  
 CC Pfam: PF00046; homeobox.1.  
 CC PRINTS: PR00024; HOMEBOX.  
 CC SMART: SM00389; HOX; 1.  
 CC PROSITE: PS00027; HOMEBOX\_1; 1.  
 CC PROSITE: PS50071; HOMEBOX\_2; 1.  
 CC Homeobox; DNA-binding; Nuclear protein; Transcription regulation.  
 KM DOMAIN 31 36 POLY-PRO.  
 FT DOMAIN 56 60 POLY-PRO.  
 FT DOMAIN 69 80 POLY-PRO.  
 FT DOMAIN 138 143 POLY-GLY.  
 FT DOMAIN 240 243 POLY-ARG.  
 FT DNA\_BIND 239 298 HOMEBOX.  
 SQ SEQUENCE 340 AA: 35855 MW: 871423225904E741 CRC64;  
 Query Match 53.8%; Score 7; DB 1; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 PPPPP 12  
 Db 69 PPLPPAP 75  
 ID NDDP\_MOUSE STANDARD; PRT; 389 AA.  
 AC 003173;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE NPC DERIVED PROLINE RICH PROTEIN 1 (NDDP-1).  
 GN NDDP1.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 CC NCBL\_TaxId=10090;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 RP TISSUE=Brain;  
 RC MEDLINE=93041923; PubMed=1420303;  
 RA Sazuka T., Tomooka Y., Katju S., Ikawa Y., Noda M., Kumar S.;  
 RT "Identification of a developmentally regulated gene in the mouse  
 RT central nervous system which encodes a novel proline rich protein."  
 RL Blochm. Biophys. Acta 1132:240-248(1992).  
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT BIOLOGICAL ROLE IN THE  
 CC DEVELOPMENT AND DIFFERENTIATION OF THE CENTRAL NERVOUS SYSTEM.  
 CC -1- TISSUE SPECIFICITY: IN THE HEART AND TESTIS AND LESS SO IN THE  
 CC LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS AND BRAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D10727; BAA01570.1; -;  
 CC MGD: MGI:97291; Nddp1.  
 CC Developmental protein.

FT DOMAIN 30 52 POLY-PRO.  
 FT DOMAIN 163 195 LEU/PRO-RICH.  
 FT DOMAIN 273 276 POLY-ARG.  
 SQ SEQUENCE 389 AA: 41249 MW: EB36C1CBEB156033 CRC64;  
 Query Match 53.8%; Score 7; DB 1; Length 389;  
 Best Local Similarity 100.0%; Pred. No. 6.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 PPPPP 10  
 Db 41 PPPPP 47  
 ID G3PT\_HUMAN STANDARD; PRT; 408 AA.  
 AC 014556; O60823;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PUTATIVE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, TESTIS-SPECIFIC  
 DE (EC 1.2.1.12) (GAPDH-2).  
 GN GAPDS.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CC NCBL\_TaxId=9606;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 RP TISSUE=Testis;  
 RC McLaughlin E.A., Hall L.;  
 RT "Nucleotide sequence of human testis-specific glyceraldehyde-3-  
 RT phosphate dehydrogenase (GAPDH-2) cDNA."  
 RL Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP Lamerdin J.E., McCreedy P.M., Adamson A.W., Burkhardt-Schultz K.,  
 RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Garmes J., Danganan L.,  
 RA Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,  
 RA Carrano A.V.;  
 RL Submitted (Jul-1997) to the EMBL/Genbank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE  
 CC + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.  
 CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
 CC DEHYDROGENASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AJ005371; CAA06501.1; -;  
 CC EMBL: AC002389; AAB64181.1; -;  
 CC InterPro: IPR00173; GAP\_DH.  
 CC InterPro: IPR002965; P\_rich\_extensn.  
 CC Pfam: PF00044; gpdh.1.  
 CC PRINTS: PR00078; G3PDHRCGNASE.  
 CC PRINTS: PR01217; PRICHEXTENSIN.  
 CC PROSITE: PS00071; GAPDH; 1.  
 KM Glycolysis; Oxidoreductase; NAD.  
 FT BINDING 224 224  
 FT ACT\_SITE 251 251  
 FT CONFLICT 220 220  
 S -> SVRAHGCPS (IN REF. 2).

SQ SEQUENCE 408 AA; 44501 MW; 301F71C768CD95D8 CRC64;

Query Match 53.8%; Score 7; DB 1; Length 408;  
Best Local Similarity 100.0%; Pred. No. 6.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
| | | | |  
Db 54 PPPPLPP 60

## RESULT 13

ALF\_PETHY STANDARD; PRT; 412 AA.  
AC 022621;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
DE ALF\_PROTEIN (ABERRANT LEAF AND FLOWER PROTEIN).  
GN ALF.  
OS Petunia hybrida (Petunia).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.  
OX NCBI\_TaxID=4102;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. W138;  
RX MEDLINE=98167899; PubMed=9435293;  
RA Souer E., van der Krol A., Kloos D., Spelt C., Bliek M., Mol J.,  
RA Koes R.;  
RT "Genetic control of branching pattern and floral identity during  
RT Petunia inflorescence development.";  
RL Development 125:733-742(1998).  
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR REQUIRED FOR THE  
CC SPECIFICATION OF FLORAL MERISTEM IDENTITY.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE FLORAL MERISTEM AND ALSO IN  
CC THE VEGETATIVE MERISTEM.  
CC -1- SIMILARITY: BELONGS TO THE FLO / LEY FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AF030171; AAC49912.1; -;  
DR Mendel; 25747; Pexx;1429;25747.  
DR InterPro: IPR002910; FLO\_LEY.  
DR Pfam: PF01698; FLO\_LEY.1.  
KW Transcription regulation; Activator; DNA-binding;  
KM Nuclear protein; Developmental protein.  
FT DOMAIN 31 44 POLY-PRO.  
FT DOMAIN 129 132 POLY-GLU.  
FT DOMAIN 169 174 POLY-GLY.  
FT DOMAIN 228 231 POLY-GLY.  
FT DOMAIN 393 397 POLY-ALA.  
SQ SEQUENCE 412 AA; 46268 MW; F1C02F8EB95AB33 CRC64;

Query Match 53.8%; Score 7; DB 1; Length 412;  
Best Local Similarity 100.0%; Pred. No. 6.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
| | | | |  
Db 35 PPPPLPP 41

RESULT 14  
ID PKNA\_MYCTU STANDARD; PRT; 431 AA.  
AC P71585;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROBABLE SERINE/THREONINE-PROTEIN KINASE PKNA (EC 2.7.1.-).  
GN PKNA OR RV0015C OR MT0018 OR MT0104.15C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=96342230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RX Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmailova M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: Z80233; CAB02435.1; -;  
DR EMBL: AE006816; AAK44240.1; -;  
DR TIGR: MT0018; -;  
DR Tuberculist; RV0015C; -;  
DR InterPro: IPR000719; Euk\_Pkinase.  
DR InterPro: IPR002290; Ser\_thr\_kin\_actsite.  
DR Pfam: PF00069; pkinase.1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE-NEG.  
DR PROSITE: PS00108; PROTEIN\_KINASE\_ST.1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM.1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
KM Complete proteome.  
FT DOMAIN 13 253 PROTEIN KINASE.  
FT NP\_BIND 19 27 ATP (BY SIMILARITY).  
FT BINDING 42 42 ATP (BY SIMILARITY).  
FT ACT\_SITE 141 141 ATP (BY SIMILARITY).  
SQ SEQUENCE 431 AA; 45597 MW; 582D183747F3C111 CRC64;

Query Match 53.8%; Score 7; DB 1; Length 431;  
Best Local Similarity 100.0%; Pred. No. 6.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10

Db 233 PPPLPP 239

RESULT: 15

GAA6\_HUMAN STANDARD; PRT; 453 AA.

AC 016445;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE GAMMA-AMINOBUTYRIC-ACID RECEPTOR ALPHA-6 SUBUNIT PRECURSOR (GABA(A)

RECEPTOR).

GN GABRA6.

OS Homo sapiens (Human);

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI\_Taxid=9606;

RN (1)

RP SEQUENCE FROM N.A.

RC TISSUE=Cerebellum;

RX MEDLINE=96226062; PubMed=8632757;

RA Hadingham K.L., Garrett E.M., Waford K.A., Bain C., Heavens R.P.,

Stinchingsinghi D.J., Whiting P.J.;

RT "Cloning of cDNAs encoding the human gamma-aminobutyric acid type A

receptor alpha 6 subunit and characterization of the pharmacology of

alpha 6-containing receptors.";

Mol. Pharmacol. 49:253-259(1996).

RL -I- FUNCTION: GABA, THE MAJOR INHIBITORY NEUROTRANSMITTER IN THE

VERTEBRATE BRAIN, MEDIATES NEURONAL INHIBITION BY BINDING TO THE

GABA/BENZODIAZEPINE RECEPTOR AND OPENING AN INTEGRAL CHLORIDE

CHANNEL.

CC -I- SUBUNIT: GENERALLY PENTAMERIC. THERE ARE FIVE TYPES OF GABA(A)

RECEPTOR CHAINS: ALPHA, BETA, GAMMA, DELTA, AND RHO.

CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -I- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

EMBL: S81944; AAB36480.1; -.

MIM: 137143; -.

DR InterPro: IPR001175; Neur\_channel.

DR Pfam: PF00065; neur\_chan; 2.

DR PRINTS: PRO0252; NRIONCHANNEL.

DR PRINTS: PRO0253; GABARECEPT.

DR PRINTS: PRO1079; GABARALPHA.

DR PROSITE: PS00236; NEUROTR\_ION\_CHANNEL; 1.

KW Postsynaptic membrane; Ionic channel; Glycoprotein; signal;

KW Transmembrane; Multigene family.

FT SIGNAL 1 19

FT CHAIN 20 453

FT DOMAIN 20 242

FT TRANSMEM 243 264

FT TRANSMEM 269 290

FT TRANSMEM 301 324

FT DOMAIN 325 419

FT TRANSMEM 420 441

FT CARBOHYD 31 31

FT CARBOHYD 128 128

FT CARBOHYD 141 141

FT DISULFD 156 170

SO SEQUENCE 453 AA: 51034 MW: 780F85845BCB107F CRC64;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 PPPLPPA 11

DB 404 PPPLPPA 410

Search completed: January 22, 2002, 16:37:08  
Job time: 318 sec

Query Match 53.8%; Score 7; DB 1; Length 453;  
Best Local Similarity 100.0%; Pred. No. 7.1;









DT 01-JUN-2001 (TEMBREl, 17, last annotation update)  
 DE ACTIVATING TRANSCRIPTION FACTOR 5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutele  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RF TISSUE=CHORIOCARCINOMA;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC005174; AA05174.1; -  
 SQ SEQUENCE 282 AA: 30690 MW: 860406631655B1 CRC64:

Query Match	69.2%;	Score 9;	DB 4;	Length 282;
Best Local Similarity	100.0%;	Pred. No. 0.082;		
Matches	9;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0

QY	4	PPPPLPAP	12
Db	131	PPPLPPAP	139

RESULT	3
Q9AA59	
ID	Q9AA59
PRELIMINARY;	
PRT;	407 AA

DT 01-JUN-2001 (Trembarel. 17, Created)  
DT 01-JUN-2001 (Trembarel. 17, Last sequence update)  
DT 01-JUN-2001 (Trembarel. 17, Last annotation update)  
DE OMPA FAMILY PROTEIN.

00 Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
0C Caulobacter.  
0X NCBI\_TaxID=69394;

RA MEDLINE=21173698; PubMed=11259647;  
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Madcock J.R.,

RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
RA Uterback T., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro B., Fraser C.M.,  
RA Deoxy R., Dooson R.O., Durkin A.S., Gwinn M.L., Hall D.N.,  
RA

KL FLOC NALLI, ACAD. SCI. U.S.A. 98:4138-4141(2001).  
DR EMBL; AE005750; AAK22732.1; -.  
TIGR; CC0747; -.  
KW Complete proteome.

Query Match	69.2%	Score 9;	DB 2;	Length 407;
Best local similarity	100.0%	Pred. No. 0	11:	

QY	4	PPPLPPAP	12
360	PPPLPPAP	368	

RESULT	4
O9FTT15	
TD COMTE	
DDRTTTHNRY.	
DEE.	
CC 11	

DT	01-MAR-2001	(TREMblrel, 16, Created)
DT	01-MAR-2001	(TREMblrel, 16, Last sequence update)
DT	01-MAR-2001	(TREMblrel, 16, Last annotation update)

•

OS *Oryza sativa* (Rice).  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta  
 OC Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae;  
 OC Erihartoideae: Oryzae;  
 OX NCBI\_TaxID=4530;

RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa r1pnpnpbare(ga3) genomic DNA, chromosome 1, PAC  
RT clone:p0410E01.";  
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AP002866; BAB17094.1;  
SO SEQUENCE 66 AA; 7421 MW; 8D53D01574E867C CRC64;

Query Match	61.5%	Score	8	DB	10	Length	66
Best Local Similarity	100.0%	Pred. NC.	0.24				
Matches	8	Conservative	0	Mismatches	0	Indels	0
						Gaps	0

QY	4	PPPLPPA	11
Db	4	PPPLPPA	11

RESULT	5
Q9VNZ6	
ID	Q9VNZ6
	PRELIMINARY;
	PRT; 108 AA

DT	01-MAY-2000	(Tremblrel. 13, Created)
DT	01-MAY-2000	(Tremblrel. 13, Last sequence update)
DT	01-JUN-2000	(Tremblrel. 14, Last annotation update)
DE	CG14571	PROTEIN.

0C Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
0C Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
0C Ephydroidea; Drosophilidae; Drosophila.

RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;

RA Amadiakues F.S., Scheier S.E., Li F.W., Roskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazel R.S., Chande M., Pfeiffer B.D.

RA Borkova D., Botchan M.R., Bouck J., Brokstein P.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhargava D., Bolshakov S.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Adili C.F., Agbayan A., Au H.-Y., Aulicews-Plamkocn C., Baldwin D.,  
 RA

RA Durbin K J, Evangelista C F, Ferraz C, Fariñeira S, Fleischmann W,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Cherry J.M., Cawley S., Danke C., Davenport L.B., Davies P.,  
RA

RA Glodex  
RA Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Telati M., Ketchum K.A., Varnon C.H., Ko Z., Konatison T.A., Ketchum K.A.

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Purl V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Slden-Klamos I., Simpson M., Skupski M.P., Smith T.,

RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.M., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003595; AAF51766.1; -  
 DR FlyBase: FBgn0037118; CG14571.  
 SQ SEQUENCE 108 AA; 11926 MW; B67F87ACB2D3A45 CRC64;

Query Match 61.5%; Score 8; DB 5; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 0.36;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 APPPLPP 10  
 Db 20 APPPLPP 27

RESULT 6  
 ID 099MA0 PRELIMINARY; PRT; 179 AA.  
 AC 099MA0:  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DE TRANSCRIPTION FACTOR MRG2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=HEART;  
 RA Zhuang D.Z., Chou Y.-T., Yang Y.-C.;  
 RT "Structural and functional conservation of MRG family in system  
 evolution.";  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF361477; AAK30622.1; -  
 SQ SEQUENCE 179 AA; 18151 MW; 7B1A819DE7865284 CRC64;

Query Match 61.5%; Score 8; DB 11; Length 179;  
 Best Local Similarity 100.0%; Pred. No. 0.54;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
 Db 116 PPPPLPPA 123

RESULT 7  
 ID 09LW68 PRELIMINARY; PRT; 195 AA.  
 AC 09LW68:  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE GB|AAC63835.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLOMBIA;  
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.,  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=COLOMBIA;  
 RX MEDLINE=20277480; PubMed=10819329;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
 RT features of the regions of 4,504,864 bp covered by sixty pl and YAC  
 RT clones.";  
 RL DNA Res. 7:131-135(2000).  
 DR EMBL: AB015474; BAB02272.1; -  
 DR EMBL: AP00733; BAB02272.1; JOINED.  
 SQ SEQUENCE 195 AA; 21434 MW; A6BA85958BEFCE23 CRC64;

Query Match 61.5%; Score 8; DB 10; Length 195;  
 Best Local Similarity 100.0%; Pred. No. 0.58;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
 Db 179 PPPPLPPA 186

RESULT 8  
 ID 09VW01 PRELIMINARY; PRT; 309 AA.  
 AC 09VW01:  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE CG13055 PROTEIN.  
 GN CG13055.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadiot E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy J., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebo J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spindler A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003528; AAF49521.1; -;  
 DR FlyBase: FBgn0036583; CG13055.  
 SQ SEQUENCE 309 AA; 33224 MW; 9DAEB67784852A93 CRC64;

Query Match 61.5%; Score 8; DB 5; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 0.84;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
 |||||  
 DB 211 PPPPLPPA 218

RESULT 9  
 O19782 PRELIMINARY; PRT; 443 AA.  
 AC O19782;

DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
 DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)  
 DE SIMILAR TO CYTOPLASMIC INTERMEDIATE FILAMENT. NCBI GI: 1213546.  
 GN F25B2.4.

OS *Caenorhabditis elegans*.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;  
 OC Rhabdilitidae; Pelodertinae; *Caenorhabditis*.  
 OX NCBI\_TaxID=6239;

RA [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Cratton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins J., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifkin L., Koopra A., Saunders D., Showkseen R.,

RA Smaildon N., Smith A., Sonnenhammer E., Staden R., Sulston J.,

RA Thiertry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*

RT *elegans*.";

RL Nature 368:32-38(1994).

RA [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Minx P., Le T.;

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: U50197; AAA91257.1; -;

DR InterPro: IPR001664; IF.

DR Pfam: PF00038; filament; 2.

SQ SEQUENCE 443 AA; 51581 MW; 2B6A54DDCB17730B CRC64;

Query Match 61.5%; Score 8; DB 5; Length 443;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 NAPPPLP 9  
 |||||  
 DB 415 NAPPPLP 422

RESULT 10  
 O62775 PRELIMINARY; PRT; 451 AA.  
 AC O62775;  
 DR O62775;

DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
 DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)  
 DE SH3 DOMAIN BINDING PROTEIN.

GN CR16.

OS *Rattus norvegicus* (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI\_TaxID=10116;

RA [1]

RP SEQUENCE FROM N.A.

RC STRAIN-FISCHER 344; TISSUE-BRAIN;

RA MEDLINE=96436481; PubMed=8839352;

RA Masters J.N., Coleman S.L., Osterburg H.H., Nichols N.R., Finch C.E.;

RT "Modulation of a novel RNA in brain neurons by glucocorticoid and

RT mineralocorticoid receptors.";

RL Neuroendocrinology 63:28-38(1996).

DR EMBL: U25281; AAA87791.1; -;

DR InterPro: IPR003124; WH2.

DR Pfam: PF02205; WH2; 1.

DR SMART: SM00246; WH2; 1.

SQ SEQUENCE 451 AA; 45307 MW; 6E2B50D1876E9CEE CRC64;

Query Match 61.5%; Score 8; DB 11; Length 451;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
 |||||  
 DB 192 PPPPLPPA 199

RESULT 11  
 O95763 PRELIMINARY; PRT; 453 AA.  
 AC O95763;

DT 01-MAY-1999 (TRENBLREL. 10, Created)

DT 01-MAY-1999 (TRENBLREL. 10, Last sequence update)

DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)

DE WUGSC:H.DJ0871815.2 PROTEIN (FRAGMENT).

OS *Homo sapiens* (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RA [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=99063792; PubMed=9847074;

RA Sulston J.E., Waterston R.;

RT "Toward a complete human genome sequence.";

RL Genome Res. 8:1097-1108(1998).

RA [2]

RP SEQUENCE FROM N.A.

RA Bourne S., Wollam C., Fielder T.;

RT "The sequence of *Homo sapiens* PAC clone DJ0871815.";

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RA [3]

RP SEQUENCE FROM N.A.

RA Waterston R.H.;

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RA [4]

RP SEQUENCE FROM N.A.

RA Waterston R.H.;

RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

RA [5]

RP SEQUENCE FROM N.A.

RA Waterston R.;

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC004912; AAD15418.1; -;

DR InterPro: IPR002965; P\_Rich\_extensn.

DR InterPro: IPR003124; WH2.

DR InterPro: IPR003882; Pistil\_extensin.

DR Pfam: PF02205; WH2; 1.

DR PRINTS; PRO1217; PRICHEXTENSIN.  
 DR PRINTS; PRO1218; PSTLEXTENSIN.  
 DR SMART; SM00246; WH2; 1.  
 FT NON\_TER  
 SQ SEQUENCE 453 AA; 46578 MW; 43E3A9C753F97D02 CRC64;

Query Match 61.5%; Score 8; DB 4; Length 453;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 APPPLPP 10  
 |||||  
 DB 261 APPPLPP 268

RESULT 12  
 ID Q920G8 PRELIMINARY; PRT; 485 AA.  
 AC Q920G8;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE SH3 DOMAIN BINDING PROTEIN.  
 GN Crl6.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 ON NCBI\_TaxID=10116;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISCHER 344; TISSUE=LIVER;  
 RA Ashman W.H., Smith J.L., Cotman S.L., Masters J.N.;  
 RT "Gene organization and differential splicing of Crl6, a novel gene  
 expressed in CNS neurons.";  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U31168; AAC99859.1; -;  
 DR EMBL; U31161; AAC99859.1; JOINED.  
 DR EMBL; U31162; AAC99859.1; JOINED.  
 DR EMBL; U31163; AAC99859.1; JOINED.  
 DR EMBL; U31164; AAC99859.1; JOINED.  
 DR EMBL; U31165; AAC99859.1; JOINED.  
 DR EMBL; U31166; AAC99859.1; JOINED.  
 DR EMBL; U31167; AAC99859.1; JOINED.  
 DR EMBL; U31159; AAC99858.1; -;  
 DR InterPro; IPR002965; P\_rich\_extensin.  
 DR InterPro; IPR003124; WH2.  
 DR InterPro; IPR003882; Pfam1\_extensin.  
 DR Pfam; PF02205; WH2; 1.  
 DR PRINTS; PRO1217; PRICHEXTENSIN.  
 DR PRINTS; PRO1218; PSTLEXTENSIN.  
 DR SMART; SM00246; WH2; 1.  
 SQ SEQUENCE 485 AA; 49380 MW; 195673E854CB37D9 CRC64;

Query Match 61.5%; Score 8; DB 11; Length 485;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 11  
 |||||  
 DB 192 PPPPLPP 199

RESULT 13  
 ID Q9FU62 PRELIMINARY; PRT; 507 AA.  
 AC Q9FU62;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE P013F10.12 PROTEIN.  
 GN P013F10.12.

OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 ON NCBI\_TaxID=4530;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone: P013F10.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP002523; BAB17066.1; -;  
 SQ SEQUENCE 507 AA; 53021 MW; 1B550AFB74164442 CRC64;

Query Match 61.5%; Score 8; DB 10; Length 507;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 APPPLPP 10  
 |||||  
 DB 404 APPPLPP 411

RESULT 14  
 ID Q9SN05 PRELIMINARY; PRT; 564 AA.  
 AC Q9SN05;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE HYPOTHETICAL 66.6 KDA PROTEIN.  
 GN F3A4.210.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 ON NCBI\_TaxID=3702;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Barques M., Collado M.C., Navarro P., Terol J., Perez-Alonso M.,  
 RA Mewes H.W., Lemcke K., Mayer K.F.X., Quelier F., Salanoubat M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL132978; CAB62121.1; -;  
 DR InterPro; IPR002965; P\_rich\_extensin.  
 DR PRINTS; PRO1217; PRICHEXTENSIN.  
 KW Hypothetical protein.  
 SQ SEQUENCE 564 AA; 66602 MW; AD6578D575E1A2BB CRC64;

Query Match 61.5%; Score 8; DB 10; Length 564;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 11  
 |||||  
 DB 51 PPPPLPP 58

RESULT 15  
 ID Q9W1X3 PRELIMINARY; PRT; 569 AA.  
 AC Q9W1X3;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE CG9848 PROTEIN.  
 GN CG9848.

OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S.D., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Bens P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003459; AAG22198.1;-;  
 DR FlyBase: FBgn0034799; CG9848.  
 SQ SEQUENCE 569 AA; 63909 MW; D1D7043B8337A5C CRC64;

Query Match 61.5%; Score 8; DB 5; Length 569;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 APPPLPP 10  
 |||||  
 DB 501 APPPLPP 508

Search completed: January 22, 2002, 16:36:45  
 Job time: 345 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:30:53 ; Search time 22.77 seconds  
(without alignments)  
12.848 Million cell updates/sec

Title: US-09-485-529-103

Perfect score: 13

Sequence: 1 LNPAPPLPPAPQ 13

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents, AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PCRTUS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53.8	15	2	US-08-769-745-31	Sequence 31, Appl
2	53.8	15	4	US-08-602-999A-356	Sequence 356, App
3	53.8	15	4	US-08-602-999A-414	Sequence 414, App
4	53.8	18	4	US-08-602-999A-320	Sequence 320, App
5	53.8	28	2	US-08-459-568-58	Sequence 58, Appl
6	53.8	28	2	US-08-399-411-58	Sequence 58, Appl
7	53.8	28	3	US-08-769-745-39	Sequence 39, Appl
8	53.8	28	3	US-08-516-859A-58	Sequence 58, Appl
9	53.8	69	4	US-08-818-112-78	Sequence 78, Appl
10	53.8	140	3	US-08-850-961-12	Sequence 12, Appl
11	53.8	141	3	US-08-850-961-10	Sequence 10, Appl
12	53.8	281	2	US-08-810-453-2	Sequence 2, Appl
13	53.8	281	3	US-08-815-190A-2	Sequence 2, Appl
14	53.8	281	4	US-09-290-640-25	Sequence 25, Appl
15	53.8	281	4	US-09-479-524-3	Sequence 3, Appl
16	53.8	281	5	PCT-US95-00362-2	Sequence 2, Appl
17	53.8	288	3	US-08-545-196B-21	Sequence 21, Appl
18	53.8	288	4	US-09-028-327-3	Sequence 3, Appl
19	53.8	311	4	US-09-179-558-66	Sequence 66, Appl
20	53.8	445	2	US-08-900-148-2	Sequence 2, Appl
21	53.8	453	1	US-08-417-330A-18	Sequence 18, Appl
22	53.8	539	3	US-08-906-360-1	Sequence 1, Appl
23	53.8	726	4	US-09-126-980-2	Sequence 2, Appl
24	53.8	726	4	US-09-476-482-2	Sequence 2, Appl
25	53.8	834	3	US-08-539-205A-6	Sequence 6, Appl
26	53.8	933	4	US-08-764-870-14	Sequence 14, Appl
27	53.8	933	4	US-08-980-115-14	Sequence 14, Appl

28	7	53.8	1876	2	US-08-609-049A-12	Sequence 12, Appl
29	7	53.8	1876	2	US-08-609-049A-28	Sequence 28, Appl
30	7	53.8	1876	4	US-09-170-996-12	Sequence 12, Appl
31	7	53.8	1876	4	US-09-170-996-28	Sequence 28, Appl
32	7	53.8	1958	1	US-07-945-283-2	Sequence 2, Appl
33	7	53.8	2205	1	US-08-093-453B-2	Sequence 2, Appl
34	6	46.2	7	1	US-08-230-047-40	Sequence 40, Appl
35	6	46.2	7	2	US-08-769-745-26	Sequence 26, Appl
36	6	46.2	7	5	PCT-US94-01840-11	Sequence 11, Appl
37	6	46.2	8	2	US-08-612-857-8	Sequence 8, Appl
38	6	46.2	10	1	US-08-230-047-12	Sequence 12, Appl
39	6	46.2	10	1	US-08-212-190A-8	Sequence 8, Appl
40	6	46.2	10	2	US-08-900-321-8	Sequence 8, Appl
41	6	46.2	10	3	US-08-899-595-8	Sequence 8, Appl
42	6	46.2	10	4	US-09-001-984C-34	Sequence 34, Appl
43	6	46.2	10	5	PCT-US95-03610-8	Sequence 8, Appl
44	6	46.2	11	1	US-08-336-343A-25	Sequence 25, Appl
45	6	46.2	11	4	US-08-652-877-22	Sequence 22, Appl

#### ALIGNMENTS

RESULT 1  
US-08-769-745-31  
; Sequence 31, Application US/08769745  
; Patent No. 5955259  
; GENERAL INFORMATION:  
; APPLICANT: Holmes, Todd C.  
; APPLICANT: Levitan, Irwin B.  
; APPLICANT: Brandeis University  
; TITLE OF INVENTION: Mechanism for the Regulation of Ion  
; FILE REFERENCE: BR096-02  
; CURRENT APPLICATION NUMBER: US/08/769,745  
; CURRENT FILING DATE: 1996-12-19  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 31  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-08-769-745-31

Query Match 53.8%; Score 7; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
Db 3 PPPPLPP 9

RESULT 2  
US-08-602-999A-356  
; Sequence 356, Application US/08602999A  
; Patent No. 6184205  
; GENERAL INFORMATION:  
; APPLICANT: SPARKS, Andrew B.  
; APPLICANT: KAY, Brian K.  
; APPLICANT: THORN, Judith M.  
; APPLICANT: QUILIAM, Lawrence A.  
; APPLICANT: DER, Channing J.  
; APPLICANT: FOWLER, Dana M.  
; APPLICANT: RIDER, James E.  
; TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
; TITLE OF INVENTION: ISOLATING AND USING SAME  
; NUMBER OF SEQUENCES: 467  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York

STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 356:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-356

Query Match 53.8%; Score 7; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLP 10  
|11111|  
Db 6 PPPPLP 12

RESULT 3  
US-08-602-999A-414  
Sequence 414, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OUILLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLKES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 414:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-414

Query Match 53.8%; Score 7; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLP 10  
|11111|  
Db 3 PPPPLP 9

RESULT 4  
US-08-602-999A-320  
Sequence 320, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OUILLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLKES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 320:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-320

Query Match 53.8%; Score 7; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 4 PPPPLP 10  
| | | | |  
DB 9 PPPPLP 15

## RESULT 5

US-08-459-568-58  
; Sequence 58, Application US/08459568  
; Patent No. 5811304  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/459,568  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,411  
; FILING DATE: 06-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-459-568-58

Query Match 53.8%; Score 7; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLP 10  
| | | | |  
DB 15 PPPPLP 21

## RESULT 6

US-08-399-411-58  
; Sequence 58, Application US/08399411  
; Patent No. 5831008  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA

ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/399,411  
; FILING DATE: 06-MAR-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 1264  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-399-411-58

Query Match 53.8%; Score 7; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLP 10  
| | | | |  
DB 15 PPPPLP 21

RESULT 7  
US-08-769-745-39  
; Sequence 39, Application US/08769745  
; Patent No. 5955259  
; GENERAL INFORMATION:  
; APPLICANT: Holmes, Todd C.  
; APPLICANT: Levitan, Irwin B.  
; APPLICANT: Brandeis University  
; TITLE OF INVENTION: Mechanism for the Regulation of Ion  
; FILE REFERENCE: BR096-02  
; CURRENT APPLICATION NUMBER: US/08/769,745  
; CURRENT FILING DATE: 1996-12-19  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 39  
; LENGTH: 28  
; TYPE: PRT  
; ORGANISM: Rat  
US-08-769-745-39

Query Match 53.8%; Score 7; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLP 10  
| | | | |  
DB 10 PPPPLP 16

RESULT 8  
US-08-516-859A-58  
; Sequence 58, Application US/08516859A  
; Patent No. 6069231  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins

NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/516,859A  
FILING DATE: 18-AUG-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/399,411  
FILING DATE: 06-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,683  
FILING DATE: 18-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 1776  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-516-859A-58

Query Match 53.8%; Score 7; DB 3; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PPPPLPP 10  
|||||  
Db 15 PPPPLPP 21

RESULT 9  
US-08-818-112-78  
Sequence 78, Application US/08818112  
Patent No. 6290969  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US//08/818,112  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 78:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 69 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-818-112-78

Query Match 53.8%; Score 7; DB 4; Length 69;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PPLPPAP 12  
|||||  
Db 23 PPLPPAP 29

RESULT 10  
US-08-850-961-12  
Sequence 12, Application US/08850961  
Patent No. 6013517  
GENERAL INFORMATION:  
APPLICANT: Respass, James G.  
APPLICANT: De Polo, Nicholas J.  
APPLICANT: Chada, Sunil  
APPLICANT: Sauter, Sybille  
APPLICANT: Bodner, Mordechai  
APPLICANT: Driver, David A.  
TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation, Intellectual Property - R440  
STREET: P.O. Box 8097  
CITY: Emeryville  
STATE: California  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/850,961  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kruse, No. 6013517man J.  
REGISTRATION NUMBER: 35,235  
REFERENCE/DOCKET NUMBER: 930049.424C4 / 1147.005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-3520  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 140 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-850-961-12

Query Match 53.8%; Score 7; DB 3; Length 140;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
1111111  
DB 105 PPPPLPP 111

## RESULT 11

US-08-850-961-10  
; Sequence 10, Application US/08850961  
; Patent No. 6013517  
; GENERAL INFORMATION:  
; APPLICANT: Respass, James G.  
; APPLICANT: De Polo, Nicholas J.  
; APPLICANT: Chada, Sunil  
; APPLICANT: Sauter, Sybille  
; APPLICANT: Bodner, Mordechai  
; APPLICANT: Driver, David A.  
; TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation, Intellectual Property - R440  
; STREET: P.O. Box 8097  
; CITY: Emeryville  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94662-8097  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/850,961  
; FILING DATE: 05-MAY-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kruse, No. 6013517man J.  
; REGISTRATION NUMBER: 35,235  
; REFERENCE/DOCKET NUMBER: 930049.424C4 / 1147.005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-3520  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 141 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-850-961-10

Query Match 53.8%; Score 7; DB 3; Length 141;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
1111111  
DB 105 PPPPLPP 111

## RESULT 12

US-08-810-453-2  
; Sequence 2, Application US/08810453  
; Patent No. 5858990  
; GENERAL INFORMATION:  
; APPLICANT: Walslh, Kenneth  
; TITLE OF INVENTION: FAS LIGAND COMPOSITIONS FOR TREATMENT OF  
; PROLIFERATIVE DISORDERS

## NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: Federal Reserve Plaza, 600 Atlantic Avenue  
CITY: Boston  
STATE: MA

COUNTRY: US  
ZIP: 02210-2211

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/810,453

## FILING DATE:

## CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Plumer, Elizabeth R.

REGISTRATION NUMBER: 36,637

REFERENCE/DOCKET NUMBER: S1237/7004

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)720-3500

TELEFAX: (617)720-2441

TELEX: 343

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 281 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-810-453-2

Query Match 53.8%; Score 7; DB 2; Length 281;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
1111111  
DB 50 PPPPLPP 56

## RESULT 13

US-08-815-190A-2  
; Sequence 2, Application US/08815190A  
; Patent No. 6046310  
; GENERAL INFORMATION:  
; APPLICANT: Queen, Cary L.  
; APPLICANT: Schneider, William P.  
; APPLICANT: Vasquez, Maximiliano  
; TITLE OF INVENTION: Fas Ligand Fusion Proteins and their  
; TITLE OF INVENTION: Uses  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/815,190A  
; FILING DATE: 11-MAR-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/614,584  
; FILING DATE: 13-MAR-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 011823-006710US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 281 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-815-190A-2

Query Match 53.8%; Score 7; DB 3; Length 281;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
|||||  
DB 50 PPPPLPP 56

RESULT 14  
US-09-290-640-25  
Sequence 25, Application US/09290640  
Patent No. 6204055  
GENERAL INFORMATION:  
APPLICANT: Dean, Nicholas M.  
APPLICANT: Marcussen, Eric G.  
TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling  
FILE REFERENCE: ISPH-0351  
CURRENT APPLICATION NUMBER: US/09/290,640  
CURRENT FILING DATE: 1999-04-12  
NUMBER OF SEQ ID NOS: 85  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 25  
LENGTH: 281  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-290-640-25

Query Match 53.8%; Score 7; DB 4; Length 281;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
|||||  
DB 50 PPPPLPP 56

RESULT 15  
US-09-479-524-3  
Sequence 3, Application US/09479524  
Patent No. 6268350  
GENERAL INFORMATION:  
APPLICANT: Barbera-Guillem, Emilio  
TITLE OF INVENTION: Polynucleotides For Inhibiting Metastasis And Tumor Cell Growth  
FILE REFERENCE: B-37  
CURRENT APPLICATION NUMBER: US/09/479,524  
CURRENT FILING DATE: 2000-01-07  
EARLIER APPLICATION NUMBER: US 09/170,948  
EARLIER FILING DATE: 1998-10-13  
EARLIER APPLICATION NUMBER: US 60/062,733  
EARLIER FILING DATE: 1997-10-23  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: WordPerfect  
SEQ ID NO 3  
LENGTH: 281  
TYPE: PRT

ORGANISM: Homo sapiens  
US-09-479-524-3

Query Match 53.8%; Score 7; DB 4; Length 281;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
|||||  
DB 50 PPPPLPP 56

Search completed: January 22, 2002, 16:30:54  
Job time: 274 sec



